

**THE  
MOTIVATIONAL  
NECESSITY OF  
DESIRE: A STUDY  
OF A NEGLECTED  
ATTITUDE**

by

David Wall

I hereby declare that the following thesis is my own, original work:

.....

David Wall 27/02/2008

**AUSTRALIAN NATIONAL**

**UNIVERSITY**

**ABSTRACT**

**THE MOTIVATIONAL**

**NECESSITY OF DESIRE: A**

**STUDY OF A NEGLECTED**

**ATTITUDE.**

by David Wall

Supervisory Committee Co-Chairs: Dr. Brian Garrett and Professor Daniel

Stoljar.

Department of Philosophy.

In both our everyday talk and our theorizing we characterise much of our psychology in terms of desires, along with beliefs. For instance, we attempt to explain peoples' actions by attributing them with particular desires, and we attempt to predict what they will do, and how they will react to certain events on the basis of what we take them to desire. In Philosophy of Mind, for example, we attempt to understand what is involved in having each kind of state, what relations they have with other phenomena, and so on. Desires, in

particular, are appealed to by a variety of fields of philosophy to explain phenomena associated with things like moral judgments, and works of fiction. What must desires be like to be able to have these various roles?

A traditional and influential approach to understanding desires characterizes them in terms of relations with motivation. I call a particular version of this approach the **Motivational Necessity of Desire (MN)**. According to this view it is necessary that if someone has a desire then they are disposed to act in ways that they believe will bring about what they desire. MN appears intuitively plausible: it is supported by familiar facts about desire and what we observe when people have desires. Also it is widely accepted, both within Philosophy of Mind and more broadly: it is presupposed by a number of important, thriving debates in other fields of philosophy. Nonetheless, a number of seemingly persuasive arguments have been made against MN. Some of these appeal to counter examples, either of particular kinds of desire, or particular kinds of individual, for which MN appears to be false, another claims that MN is incompatible with certain kinds of empirical evidence about desires, while others argue that MN is uninformative so it is not the a relation that helps achieve an understanding of desire.

In this dissertation I attempt to defend MN against what I take to be the strongest of these objections. After briefly motivating the view by highlighting its theoretical and explanatory advantages I discuss in turn each of



these main objections in detail. I argue that in each case the objection is either based on a misunderstanding of the kind of claim that is being made about desire by MN, or on a misunderstanding of precisely what this thesis entails. By making this clear, and by clarifying the precise complaint that is being made in each objection I attempt to show that each of them fails. But they do so for interesting, and different reasons that are revealing about other properties and relations of desires. I conclude that these objections give us no reason to reject this intuitively plausible thesis about desire.

## ACKNOWLEDGEMENTS.

Many people have helped me to write this dissertation and I would like to take the opportunity to thank them here.

First, as anyone who travels overseas to study will understand, it makes a great difference if people are friendly to you and you have some kind of social life. In that respect I have been very fortunate to study at the ANU: the philosophy program here must be as socially active as any you might find elsewhere and the staff and postgraduates have made me feel welcome into that. I'd especially like to express my thanks to those I have perhaps spent most time with when away from my desk, including Daniel and Annabel, Carl, Kelly, Martin, Hatha, Philippe, Various Bens, John, Karen, Yuri, and the other, non-philosophy, students at Graduate and University Houses that I share a table with most Wednesdays. I would also like to thank the Master of University House and the staff at University and Graduate Houses who have shown me a great deal of kindness.

The philosophical environment at the ANU is incredible active and impressive and I have benefited enormously just from being involved in and exposed to this. In direct relation to my own work, both Peter Roeper and Kim Sterelny were part of my supervisory panel for short periods at different times and I appreciated their comments on my work. From amongst the other

faculty members I would like to thank Havi Carel, David Chalmers, Alan Hajek, Peter Godfrey-Smith and Jason Grossman who have been kind enough to read or discuss my work. A number of visitors to the program have also helped me in this way, including Tim Bayne, Michael Brady, Christophe Fehige, and Fiona Macpherson and I am very grateful for that. Timothy Schroeder was very generous and enthusiastic to discuss my work and his own in our extensive email correspondence and I appreciated that very much. And special thanks to Daniel Friedrich and Martin Smith from amongst the postgraduates for numerous long and helpful discussions and for reading and commenting on my work.

There are a few people to whom I am especially indebted: first, Jeremy Shearmur, the head of the Philosophy Department in the School of Humanities. Jeremy has gone far beyond what it would be reasonable to expect of someone in that position. He has always taken an interest in my work and has been endlessly encouraging. He has also shown enormous faith in my ability, in particular in offering me fantastic opportunities to gain teaching experience in the department that I have benefited from a great deal. But most of all he and Pam have been kind and supported me in so many ways. I will always be grateful for this and I hope one day to be able to repay them in some way.

My greatest academic debts are to the Co-Chairs of my supervisory panel, Brian Garrett and Daniel Stoljar. I worked almost exclusively with Brian for the first 2 years of the degree. Brian was as generous with his time as you could hope for from a supervisor, meeting with me as regularly as I could write things for him to read, and always giving me detailed feedback.

Daniel is perhaps the person most responsible for my completing this dissertation and I cannot overstate his contribution. Not only did he go out of his way to keep in contact with me while I was applying to come to ANU from the UK, and so to actually get me here, but in the past 18 months in particular what I have learnt from him has almost completely changed how I do philosophy. The particular direction for the dissertation came from him and he has discussed every version of every part of it with me. I have tried to reflect his influence throughout and I think that the most successful parts are those where I have achieved this to the greatest extent.

Finally, I want to thank my (complicated) family for their support, especially my mother: my failure to act in the ways that she would like is not due to my lacking the dispositions to do so, rather it is due to my lacking the means-ends beliefs.

# CONTENTS

Abstract of the Thesis .....	
Acknowledgments .....	i
Contents.....	iv
Introduction .....	1
Desire.....	1
Structure of the Thesis.....	24
Chapter 1 – Motivation for Motivational Necessity .....	30
1.1 – Introduction .....	30
1.2 – MN in Different Debates .....	31
1.3 – MN in Philosophy of Mind .....	39
1.4 – A Fact about Desire .....	49
1.5 – Explaining the Fact.....	53
1.6 – Conclusion.....	61
Chapter 2 – Motivation, Desires and the Weather Watchers. ....	63
2.1 – Introduction.....	63

2.2 – Strawson’s Objection.....	64
2.3 – Options .....	68
2.4 – Two Versions of the Weather Watchers .....	75
2.5 – Wayne the Weather Watcher.....	78
2.6 – Shane the Weather Watcher .....	80
2.7 – Responses to Shane .....	82
2.8 – Motivation for Shane.....	85
2.9 – A Regress of Intentions and Actions.....	88
2.10 – Conclusion.....	91
 Chapter 3 – Troublesome Desires .....	 93
3.1 – Introduction .....	93
3.2 – Self Passive Desires .....	95
3.3 – Options .....	97
3.4 – Self Passive Desires are not a Counter Example to MN .....	101
3.5 – Self Passive Desires and Straightforward Desires .....	105
3.6 – Options (part 2).....	106
3.7 – Two Dispositions .....	110
3.8 – Distinguishing Between Desires.....	114
3.9 – Necessity and Impossibility Desires .....	115
3.10 – Options (part 3).....	119

3.11 – Conclusion.....	124
 Chapter 4 – Defending Dispositional Direction of Fit.....	126
4.1 – Introduction.....	126
4.2 – The Explanatory Power of MN.....	127
4.3 – Direction of Fit .....	130
4.4 – Counter Examples to Smith’s DOF Account.....	137
4.5 – Sobel & Copp’s Uninformativeness Objection .....	144
4.6 – Responses to Sobel & Copp.....	147
4.7 – Humberstone’s Uninformativeness Objection.....	150
4.8 – Responses to Humberstone.....	157
4.9 – Relevance to MN.....	160
4.10 – Normative Objections to Smith’s DOF account.....	163
4.11 – Conclusion.....	165
 Chapter 5 – Desire and Reward.....	167
5.1 – Introduction.....	167
5.2 – Two Functionalist Methodologies.....	169
5.3 – An Information Condition on a Theory of Desire.....	177
5.4 – The Natural Kinds Objection.....	181
5.5 – Schroeder’s Response to the Natural Kinds Objection .....	187

5.6 – Schroeder’s Reward Theory of Desire .....	190
5.7 – Alternative Responses to the Natural Kinds Objection .....	199
5.6 – The Reward Theory of Desire and MN.....	210
5.7 – Conclusion.....	218
Conclusion .....	221
Bibliography.....	246



# INTRODUCTION

## **Desire.**

Desire plays a fundamental role in our lives: Spinoza called it ‘the very essence of man’ (Spinoza 1677, Ethics). Samuel Johnson agreed, saying that ‘some desire is necessary to keep life in motion, and he whose real wants are satisfied must admit those of fancy’ (Johnson 1759, The History of Rasselas). We all know what it is like to desire something, the range of objects that people desire seems almost without limits, and people often go to tremendous lengths in an attempt to satisfy their strongest desires. They may even re-structure their entire lives to do so. We all know what it is like both to have a desire satisfied and to have a desire frustrated, and we mainly prefer the former to the latter.

Despite their importance we do not think that our desires should go unconstrained, that we should let desiring and satisfying our desires be the sole aim of our lives. For example, Cicero advised people to ‘let your desires be ruled by reason’, while Lao-tzu, the founder of Taoism, blamed the fact that people desire for the problems faced by individuals and societies: he said that ‘without desire there would be tranquility. In this way all things would be at peace’.

But what are desires? What are we saying about Amy if we say that she wants a cup of coffee? What do Bobby and Claire have in common if Bobby desires a sports car and Claire desires that there is world peace? How do Dean and Emily differ if Dean wants to drown his sorrows while Emily wants her Daddy to buy her a pony? How can Frances be the object of Graham's desires like a bar of chocolate is the object of Harriet's?

One influential and intuitive view is that desire has an essential connection with motivation: that necessarily having a desire motivates someone, *ceteris paribus*, to act to try to satisfy their desire. I call a particular version of this view the Motivational Necessity of Desire, or **Motivational Necessity (MN)** for short. According to MN, for example, if Isobel wants a cup of coffee she is motivated to do things like go to the kitchen, boil the kettle and make a cup, or go to the café, place her order with the barista and wait in line to collect her drink. And if Jim wants Kerry to marry him he is motivated, for example, to take her to a romantic restaurant, get down on one knee and present her with a ring. I will specify MN more precisely below, and my aim in this dissertation is to defend it against a number of seemingly persuasive objections.

The view that desire is necessarily connected with motivation has a long tradition in philosophy. Hume, for example, said that desires are 'motivating passions' (Hume 1739/1965). Similarly, Locke said that desire is

‘an *uneasiness* of the Mind for want of some absent good’ (Locke 1689/1975, p251). According to Locke all motivation to act in a particular way comes from having this kind of uneasiness, and if someone is unmotivated then this is because they lack such an uneasiness. He said,

[t]he motive, for continuing in the same State or Action, is only the present satisfaction in it; The motive to change, is always some uneasiness: nothing setting us upon the change of State, or upon any new Action, but some uneasiness. (Locke 1689/1975, p249).

More recently, in the 1940’s and 50’s Behaviourists like Skinner (1953) and Ryle (1949) attempted to analyse all kinds of mental states, including desires, in terms of patterns of actions and dispositions to act. According to Ryle to have a desire, or an emotion or some similar kind of state is ‘to tend to act in [...] innumerable [characteristic and related] ways’ (Ryle 1949, pp. 83-84). He said that attributions of these states are attributions of dispositional properties: these statements are ‘elliptical expressions of general hypothetical propositions of a certain sort’ (Ryle 1949, p83), namely that someone will act in particular ways in particular circumstances.

In current philosophy the view is prominent too. Functionalists about mental states, who attempt to analyse particular mental states in terms of their

relations, commonly claim that among the relations that characterise desire is a connection with action.<sup>1</sup> Stalnaker, for instance says that, '[t]o desire that P is to be disposed to act in ways that would tend to bring it about that P in a world in which one's beliefs, whatever they are, were true' (Stalnaker 1984, p15). Similarly, Michael Smith says that,

[w]e should think of desiring to  $\varphi$  as having a certain set of dispositions, the disposition to  $\psi$  in conditions C, the disposition to  $\_$  in conditional C $\_$ , where in order for conditions C and C $\_$  to obtain, the subject must have, inter alia, certain other desires, and also certain means-ends beliefs, beliefs concerning  $\varphi$ -ing by  $\psi$ -ing,  $\varphi$ -ing by  $\_$ -ing and so on (Smith 1994, p112).

However, this view that desire and motivation are necessarily connected is not accepted by everyone. Some of those who offer accounts of desire emphasize different properties and relations. Some, appeal to connections between desire and experiences of pleasure:<sup>2</sup> Aristotle, for example, called desire 'the craving for the pleasant', while Mill said that 'desiring a thing and finding it pleasant [are] two modes of naming the same

---

<sup>1</sup> See for example, Braddon-Mitchell and Jackson (1996), Humberstone (1992), Lewis (1983), Smith (1994), Stalnaker (1984), and so on.

psychological fact' (Mill 1863). Others have focused on the connection between desiring something and evaluating it favourably.<sup>3</sup> And some people have explicitly denied that desire is necessarily connected with motivation. Some of them highlight cases of extreme depression, accidie and disillusionment where people appear to lack all motivation despite seeming to have desires.<sup>4</sup> They also highlight cases of akrasia, in which someone appears to act in a way that is in conflict with obtaining what they desire.

As I understand it there are six main objections that have been made to this view. Of these, four are based on various kinds of counter example: they present cases in which a subject has a desire but seems to lack the kind of motivation that they would have if desire and motivation were necessarily connected in the way described by MN. Another objection argues that this connection is incompatible with relevant empirical information: it is argued that best current theories in neuroscience suggest that there is no necessary connection between having the brain structures that are responsible for having desires and having those that are responsible for being motivated to act. A further objection argues that the view is uninformative as opposed to being false. Because they think that it does not say anything interesting about desire

---

<sup>2</sup> See for example, Aristotle in *De Anima*, Mill (1863), Fehige (2001), and Kim (1998) among others.

<sup>3</sup> See, for example, Anscombe (1957), Platts (1979), Oddie (2005), Stampe (1987), and so on.

<sup>4</sup> Such as those described by Stocker (1979) for example.

the people who make this objection deny that desire should be analysed in terms of a connection with motivation.

In this dissertation I offer a limited defense of this view, that desire and motivation are necessarily connected. In the following chapters I will consider each of these main objections in turn and argue that it fails to show that a particular version of the view is false or that it is uninformative. The thesis I will defend, and that I will hereafter mean by Motivational Necessity (MN) is the following:

(MN)            If *S* desires that *p* then *S* is disposed to act in ways *S* believes will bring about *p*.

Notice a number of important features of MN: first, MN is a claim about a necessary condition for having a desire. It is not a claim about a sufficient condition. It is clear then that MN is not a theory or complete analysis of having a desire, if we take a theory of something to be a set of necessary and sufficient conditions for having that thing. MN could be part of a theory of desire, most straightforwardly in a theory that identified having a desire with having this kind of disposition to act. But MN is not such a theory: it could equally be part of a more complex theory of desire in combination with other necessary conditions that together are said to be sufficient for having a desire.

We might call any such theory of desire, that has MN as a necessary condition, a version of ‘Motivational Theory of Desire’.

Second, MN is a claim about desire in particular. This is the kind of state that we attribute when we say things like ‘Lionel has a desire to go to the Caribbean on holiday’, and ‘Marie wants a cup of coffee’. There are a number of kinds of pro attitudes, like desire, hope, wish and intention: these are sometimes together called ‘conative states’ to distinguish them from states like belief, assumption, imagination, supposition, sometimes together called ‘cognitive states’. I will not attempt to give an account of what distinguishes one kind of conative state from the others. Rather, I take it that there is intuitively a difference between say, the mental state that someone has when they desire to have some chocolate and the mental state that someone has when they wish to have some chocolate. MN is only a thesis about one of these pro-attitudes or conative states: desire.

Third, as I understand them desires are states that have an object. That is, when someone has a particular desire there is something that they desire to have. In this respect, desires are like beliefs and unlike sensations, for example. When someone has a belief there is something that is the object of their belief. In contrast, when someone has a sensation of pain there does not seem to be something that is the object of their pain. But what kind of thing are the objects of desires?

In the case of beliefs it is generally assumed that the object is a proposition, where propositions are intentional objects of some kind and are what is expressed in the 'that\_' clause of an intentional expression. For example, when Frank believes that Canberra is East of Sydney, the proposition that is the object of his belief is the proposition that *Canberra is East of Sydney*. There are good philosophical reasons in support of this view about the objects of belief. These are the familiar reasons for thinking that the object of belief cannot be some concrete object or state of affairs in the world: for example, when someone has a belief about a non-existent thing, like a belief that Santa Claus has a white beard, or if they have a belief that is false then the object of their belief cannot be the concrete object or state of affairs that is described by the belief: this state of affairs does not obtain in the world. Similarly, if someone has a belief that such-and-such does not exist, and this belief is true, then the object of the belief cannot be something that exists in the world. In addition, if someone has a belief like the belief that some dogs have three legs then it seems that there is no particular thing, no particular object or state of affairs that their belief is about<sup>5</sup>. Rather, it is a belief about dogs in general. Admittedly there are contexts in which we attribute beliefs that seem not to have propositions as their objects, when we say things like

---

<sup>5</sup> This is like the problem, described by Locke, about how we can have think about general things, like a thought about man, but not about any particular man, or a thought about all triangles, where the triangle we are thinking about is neither isosceles, nor equilateral, nor scalene.



‘Nancy believes Oliver’, or ‘Peter believes in the Ten Commandments’. But it is generally assumed that these can be paraphrased to the propositional form without changing the meaning of the belief attribution: in these cases as something like ‘Nancy believes that what Oliver said was true’, and ‘Peter believes that people should obey the Ten Commandments’ respectively.

As I understand it, the common consensus is that the objects of desires are propositions like the objects of beliefs.<sup>6</sup> The reasons to accept this view about the objects of belief seem to apply to the objects of desire as well, *mutatis mutandis*. Again there are contexts in which we attribute desires in a non-propositional form, for example when we say ‘Queenie wants a bar of chocolate’, or ‘Roger has a desire to go swimming’. It seems more difficult to paraphrase these cases into the propositional form with precision than in the analogous cases of attributions of non-propositional beliefs. However, I will assume that it is possible to do this and, following the consensus I will assume that the objects of desires are propositions. Of course, there are complicated issues concerning what propositions are and what the relationship is between a mental state and its propositional object. However, such issues are not my concern in this dissertation.<sup>7</sup>

---

<sup>6</sup> See for example Field (1978), Smith (1994), Stalnaker (1984), Stampe (1987). See also Strawson who uses takes his argument against MN to generalize to states like beliefs because both desires and beliefs have in common that their object is a proposition. He says, ‘Chapter 9 considers the case of desire and hence, more generally, the case of the so-called “propositional attitudes”’ (Strawson 1998, p434).

<sup>7</sup> See for example Field (1978) and Stalnaker (1984, Chapter 1) for more detailed discussion.

A fourth feature to notice about MN is that it claims that having a desire is necessarily connected with having a disposition to act. Contrast it with a thesis that might be called the Action Necessity of Desire (AN) that claims having a desire is necessarily connected in the same way with actually acting. AN says the following:

(AN)            If *S* desires that *p* then *S* will act in a way *S* believes  
                         will bring about *p*.

MN is not like this. Of course, MN says that having a desire is connected in some way with actually acting. But this is in terms of the relation between being disposed to act and actually acting.

But what does it mean to say that something has a particular disposition? Consider the case of solubility, a paradigm example of a disposition. To say that sugar is soluble in water is not to say that in all cases in which a sample of sugar is placed in water that it will dissolve. Rather it is to say that if a sample of sugar is placed in water and a particular set of conditions obtain then the sugar will dissolve. These might be called the ‘manifestation conditions’, or ‘enabling conditions’ for the disposition: they can be thought of as a set of propositions that must be true for a particular sample of sugar to dissolve if it is placed in water. For example, to say that

sugar is disposed to dissolve in water is to say that if a particular sample of sugar is placed in water and  $x, y, z \dots$  and  $n$  are true then the sugar will dissolve. It is a notoriously difficult task to specify what this list of propositions is for any particular disposition: there are complicated issues about how to specify them so as to avoid cases of interference like masking or finkishness, where intuitively the object has the disposition despite not manifesting it when we would expect. However, these issues are not something that I will address in this thesis.<sup>8</sup>

I understand MN as making an analogous claim about the connection between desire and acting. That is, it could be expanded in the following way:

If  $S$  desires that  $p$ , and  $x, y, z$  and  $\dots$  and  $n$  are true then  $S$  will act in a way  $S$  believes will bring about  $p$ .

In chapter 2 I will discuss again the problem of specifying which propositions are in the set that make up the manifestation conditions of a disposition to act:

---

<sup>8</sup> See for example, Lewis (1973), Fara (2005), for more detailed discussion. Masking occurs when the external conditions are interfered with to prevent a disposition from manifesting when it normally would. For example if a vase were wrapped in bubble-wrap before it was dropped over a hard surface then it wouldn't break. Nonetheless it is intuitive the vase is still fragile: its disposition to break when dropped has been masked. Finkish interference occurs when the internal structure of the object that intuitively has a particular disposition is altered just in those cases in which it would normally manifest: for example if a sorcerer magically changed a china vase into concrete just in cases in which it was dropped over a hard surface then again it would not break. But again it is intuitive that the vase is still fragile. Its disposition to break when dropped has been finkishly blocked.

in particular I will discuss whether having an intention to act is a condition of manifesting a disposition to act.

What about the particular disposition that MN is concerned with, the disposition to act? For the purposes of this dissertation I take having a disposition to act to mean having a disposition to perform certain bodily movements. So acting entails moving your body. However, moving your body does not entail acting: someone's actions are only a subset of their bodily movements. The important distinction here is the intuitive one between those movements that are things that you do and those that are merely things that happen to you. This is the distinction that Davidson attempts to explicate in his 1971 essay 'Agency'<sup>9</sup> when he asks,

[w]hat events in the life of a person reveal agency; what are his deeds and his doings in contrast to mere happenings in his history; what is the mark that distinguishes his actions? (Davidson 1980, p43).

Intuitively there is a genuine difference between those bodily movements that are someone's actions and those that are mere bodily movements. For example, a case in which someone acts to straighten their leg at the knee is

---

<sup>9</sup> Davidson, D.: 1971: 'Agency', in Agent, Action and Reason [Binkley, R., Bronaugh, R., and Marras, A. (eds.)] (University of Toronto Press). Reprinted in Davidson, D.: 1980: Essays on Actions and Events (OUP: New York).

different from a case in which it straightens merely as a reflex, because their patella tendon is struck: and these cases are different in certain real respects and not merely in the way that they are described. According to Davidson the difference is that the bodily movements that are a person's actions are those movements that they do intentionally whereas the mere bodily movements are those movements that are not done intentionally:

Consider coffee spilling again. I am the agent if I spill the coffee meaning to spill [it], but not if you jiggle my hand. What is the difference? The difference seems to lie in the fact that in one case, but not the other, I am intentionally doing *something*. My spilling the contents of my cup was intentional (Davidson 1980, pp. 45-46).

I will follow Davidson here. So according to MN if someone has a desire then they are disposed to intentionally move their body in a particular way. I will discuss this further in chapter 2. There are other complicated issues here concerning deviant causal chains for example, but I will not attempt to address those in this dissertation at all.<sup>10 11</sup>

---

<sup>10</sup> See for example, Davidson, D. (1980): 'Freedom to Act'; Essays on Actions and Events (OUP: New York) for detailed discussion.

<sup>11</sup> Does this assumption, that acting entails moving your body, rule out the possibility of disembodied minds? In fact it does not do this: Firstly all that it entails at most is that disembodied minds can have desires. It entails nothing about other kinds of psychological states. So it is consistent with there being a disembodied mind that can have beliefs,

Another feature to note about MN concerns the particular actions that it says someone is disposed to perform if they have a desire: these are actions that are related to their beliefs about how to bring about what they desire. So it is consistent with their actions failing to bring about what they desire: something that might occur if their beliefs about which actions they can perform to do this are false for example. Contrast MN with a different thesis, call it ‘MN(success)’, that claims that having a desire entails having a disposition to act in ways that will succeed in bringing about what you desire. MN(success) says the following:

[MN(success)] if *S* desires that *p* then *S* is disposed to act in ways that  
will bring about *p*.

---

sensations and so on. Second, it is not inconsistent with things like ghosts and angels having desires except where having a body is understood as having something with bulk, that is something that is physical, has mass, extension, and so on. There are people who take having a body to imply this: indeed some of these people take this to show that the disembodiment is incoherent, and that things like ghosts and angels are not possible [see for example Shoemaker, S. 1996: *The First-Person Perspective and Other Essays*, (Cambridge University Press, Cambridge)]. However, MN is not committed to that. MN merely requires someone to have something that it can move for it to be able to have desires. So MN is consistent with there being angels and ghosts, like Casper the friendly ghost, or the ghost of Jacob Marley in Dickens’s *A Christmas Carol* having desires. These things have a body in the sense of having something that they can move: Marley walks up the stairs in Scrooge’s house for example. So MN is consistent with these kinds of thing having desires. Third, there are some extreme cases, like a mere locus of thought, that MN entails cannot have desires. But this kind of extreme case creates problems for a wide range of otherwise plausible views: for example, they are a problem for causal theories of knowledge if they are supposed to lack any physical properties. They are a problem for various kinds of externalism about meaning: if they cannot enter into causal relationships with anything in the world then what are their thoughts about? The fact that MN is inconsistent with such extreme cases having desires is not obviously a special problem for the view.

MN(success) entails that if  $S$  desires that  $p$  and, as a matter of fact, doing  $\varphi$  is the only action that  $S$  can perform that will bring about  $p$  then  $S$  is disposed to do  $\varphi$ . Even if  $S$  does not believe this about  $\varphi$ , or  $S$  believes that doing  $\varphi$  would not bring about  $p$  then  $S$  is disposed to do  $\varphi$  nonetheless. According to MN(success), how someone is disposed to act if they have a particular desire is independent of their beliefs. MN(success) seems implausible, but MN is not like this: rather, according to MN the actions someone is disposed to perform if they have a particular desire depends on their beliefs about what they can do to satisfy their desire. For example, if Sarah believes that she can bring about world peace by picketing the White House then according to MN, if she has a desire that there is world peace, she is disposed to picket the White House. The beliefs that are relevant here are beliefs like a belief that by doing  $\varphi$  I can bring about  $p$ , what I will call ‘instrumental beliefs’, or ‘means-ends beliefs’. Someone might have more than one instrumental belief about different actions they can perform that would satisfy their desire, in which case according to MN they are disposed to act in more than one way: for example, Sarah might also believe that she can bring about world peace by writing letters to the UN, in which case if she has a desire that there is world peace she is disposed to write letters to the UN and she is disposed to picket the White House. On the other hand, someone might lack any instrumental beliefs that are relevant to their particular desire: it is possible that over the course of time Sarah comes to

stop believing that she can bring about world peace by picketing the White House or by writing letters to the UN and forms no new beliefs about how she might bring about world peace. But suppose that she maintains her desire that there is world peace. This case might appear to be a problem for MN but in fact it is not. It is true that while Sarah continues to lack any appropriate instrumental beliefs she will not act on her desire that there is world peace. Nonetheless, she is still disposed to act to try to bring about world peace because if she regains her earlier beliefs or forms different instrumental beliefs then, if certain other conditions obtain she will act on this desire. What this suggests is that having an appropriate instrumental belief is among the manifestation conditions of the disposition to act that MN claims is necessarily connected with having a desire. But this should seem plausible given the way that MN differs from MN(success) with respect to the relation between a person's beliefs and the particular actions that MN says they are disposed to perform.

People sometimes talk about desires being 'satisfied' and 'unsatisfied', but it is often unclear what is meant by these terms in different contexts. For instance, someone might claim that having a desire satisfied involves becoming aware in some way that what you desire has come about: that is, roughly, that if  $S$ 's desire that  $p$  has been satisfied then  $S$  is aware that  $p$  is the case. Others might claim something stronger than this, having a desire



satisfied requires not only becoming aware in some way that what you desire has come about but also having some kind of positive affect or feeling as a result of having this awareness: that is, roughly, that if  $S$ 's desire that  $p$  is satisfied then  $S$  is aware that  $p$  and  $S$  has a positive affect (caused in the right way by  $S$ 's awareness that  $p$ ). This raises interesting issues, for example about the possible objects of desire, about whether desires can persist, or continue to be held following their satisfaction, about self-knowledge, how we come to know what it is we desire, what kinds of cognitive access and what kinds of positive affects or feelings are related to desires, about what the relation between desires and aversions is, about the relation between desires being satisfied and their being unsatisfied is, and so on. However, I will attempt to avoid questions like these here: for the purposes of this dissertation any reference to a desire being satisfied, or someone acting to try to satisfy their desire should be understood merely in terms of what is desired being the case. I take the relation between satisfaction and desire to be analogous to the relation between truth and belief. Just as  $S$ 's belief that  $p$  is true iff it is the case that  $p$ , so  $S$ 's desire that  $p$  is satisfied iff it is the case that  $p$ . And  $S$ 's desire that  $p$  is unsatisfied iff it is not the case that  $p$ . So having one's desire satisfied does not entail anything about one's awareness of this or any affective response one might have to one's desire being satisfied.

One last feature of MN that I will highlight here is the kind of relation that MN is concerned with. According to MN if someone has a desire then there is a certain fact about them: that is, that they have a disposition to act in a particular way. This might consist in their having a certain natural feature, like a particular intrinsic property for example, that will cause them to move their body in a particular way if certain other conditions obtain. However, the thesis is not committed to any particular claim like this. Moreover, MN is not a claim about any normative relations of desire. Contrast MN with a different thesis about a relation between having a desire and what it is rational to do, call it ‘MN(rat)’. MN(rat) says the following:

[MN(rat)]      If  $S$  desires that  $p$  then  $S$  is disposed to act in ways  $S$  believes will bring about  $p$ , insofar as  $S$  is rational.

MN(rat) is a weaker thesis than MN: it says that MN is true of the rational subjects. Whereas MN says that it is true of all things that if they have a desire then they have a disposition to act, MN(rat) restricts this claim to a subset of all things, that is the things that are rational. So MN(rat) is entailed by MN and objections to MN(rat) will be objections to MN. However, for the purposes of this dissertation I will focus on the unrestricted thesis, MN.

This is not to deny that there are normative facts about desire. It is commonly thought that having a particular desire can have implications about what it is rational for someone to do, for example that if someone desires that  $p$  then, in virtue of this they have a reason to act in ways that they believe will bring about  $p$ .<sup>12</sup> And it is often thought that there are facts about the kind of objects that it is appropriate to desire, for example that something is only desirable if it has value, either in virtue of having some valuable characteristic or as a means to achieve something with a valuable characteristic.<sup>13</sup> These are interesting issues: however, I will not address any normative features of desire like these in this dissertation.

I take it that my project here, to defend MN is not a trivial matter. Earlier we saw how prominent and important desire is in our personal lives. In addition desires are significant within many fields of philosophy. So there are philosophical reasons to want a better understanding of desire as well as personal reasons. Their most obvious relevance is within Philosophy of Mind. Many people working in Philosophy of Mind attempt to provide analyses of different kinds of mental state, that is give necessary and sufficient conditions for having that kind of state. People do this for belief, perception, intention,

---

<sup>12</sup> See for example Stampe (1987) for discussion.

<sup>13</sup> See for example Anscombe (1957) for discussion.

emotion and so on.<sup>14</sup> In light of the prominence of desire it is not unreasonable to expect some philosophers of mind to be trying to give the same kind of analysis of them.<sup>15</sup> But the notion of desire is also used extensively in other, less obvious fields, such as Meta-Ethics, Theories of Practical Reasoning, Aesthetics and Philosophy of Language, among others.<sup>16</sup>

In Meta-Ethics, for example in the debate between Internalists and Externalists, in particular where this intersects with the debate between Descriptivists and Expressivists, certain positions in the debate appeal to desires to explain the practical character of moral judgments. People who hold certain versions of Expressivism and those who hold certain versions of Internalism claim that making a moral judgment must consist in or entail having a desire to act in accordance with that moral judgment: that is that if someone judges that the right thing to do is to  $\varphi$  then this consists in or entails that they have a desire to  $\varphi$ . These Expressivists and Internalists say that this

---

<sup>14</sup> See for example, Davidson (1980) and Bratman (1984) for analyses of intentions, Prinz (2005) for an analysis of emotions, etc.

<sup>15</sup> As, indeed a number of people have attempted to do, and in a variety of ways, as I discuss in the next Chapter, in section 1.5. However, as I understand the field, the number of people who work on this kind of project, and the extent to which it is discussed and has been developed is significantly less than for the projects of giving analyses of other kinds of mental state.

<sup>16</sup> However, caution must be taken here. Merely because a field of philosophy uses the term 'desire' it should not be assumed that it is using it in the same sense as we do when we talk about the mental state of desiring. So it should not be simply assumed that the findings from an analysis of desire in philosophy of mind will be straightforwardly applicable to other fields in which the notion of desire is used. Nonetheless, it is perhaps reasonable to assume that there will be something in common between the notions used in other fields and desire as a mental state. If the relationship can be understood then it may be possible for an analysis of desire as a mental state to be relevant elsewhere.

is the only way to account for the fact that people seem to be motivated to act in accordance with their moral judgments.

In Practical Reasoning, for example in an analogous debate between Internalists and Externalists, various positions in the debate appeal to desires to explain the practical character of judging that you have a reason to act in a particular way. Such Internalists claim that when someone judges that they have a reason to  $\varphi$  then *ceteris paribus* this entails that they have a desire to  $\varphi$ . Analogously, they take it that this is the only way to account for the fact that we seem to be motivated by our judgments that we have a reason to act in a particular way.<sup>17</sup>

In Aesthetics and Philosophy of Language, for example when attempting to explain our engagement with and reactions to fiction, some theorists appeal to desires when they claim that fictions are props for games of make-believe or pretense. Roughly, people who propose to analyse fictions in terms of pretense claim that works of fiction encourage people to engage in certain kinds of imagining and they provide the rules for how such imaginings should proceed and how it is appropriate to react [see for example, Walton (1990), Currie (2004)]. One of the debates here between different versions of

---

<sup>17</sup> Interestingly, in light of the fact that many take the case of moral judgments to be just a particular instance of this kind of practical judgment, where we judge that the reason we have to act in a particular way is a moral reason, there is less explicit debate about expressivist and descriptivist positions about reason judgments [although see e.g. Gibbard (1990). Blackburn (1984) also has expressivist views about statements in other areas of discourse, such as modality and causation].

this account is whether our reactions to fictions involve us having genuine desires, say to behave in a way that is similar to how someone would behave in the imagined situations [see for example, Nichols and Stich (2003)] or whether these reactions involve us having another kind of state that is only desire-like but that is not a genuine desire [see for example, Currie and Ravenscroft (2002), Velleman (2000)].

Perhaps surprisingly, in light of this personal and philosophical importance, desire has been under-researched in its own right. In contrast to the other kinds of mental state that we most commonly have, states like belief, perception, intention, and emotions, there has been little recent study of desire as a mental state. For example, there are reasonably well established views about belief and we take ourselves to have a good understanding of many of its important features.<sup>18</sup> In addition there has been considerable discussion of, and detailed attempts to explain, various puzzles about belief.<sup>19</sup>

The other common kinds of mental state have also received a significant amount of attention in recent philosophy: for example, there is a

---

<sup>18</sup> For example, for work on the propositional objects of belief see e.g. Adams (1987), Chalmers (1996), (2000), (2006), Chisholm (1982), (1984), (1989), Frege (1879), Jackson (1988), (1996), Kaplan (1978), Kripke (1980), Lewis (1979), (1986), Perry (1979), Pollock (1982), Russell (1903), (1910), Sorensen (1988) (2002), Tienison (1984), Van Inwagen (2003), etc, for work on the metaphysics of belief see e.g. Baker (1987), (1993), Cohen (1996), Dennett (1983), Garfield (1988), Sobel and Copp (2001), Schwitzgebel (2002), van Gulick (1994), etc.

<sup>19</sup> For example, for work on the lottery and preface paradoxes, in which people appear to have inconsistent beliefs see e.g. Appiah (1987), Cohen (2005), Haack (1996), Sorensen (1988), Weintraub (2001), Wheeler (2005), Williamson (1996), etc, for work on delusional beliefs

large and rapidly growing literature on the emotions.<sup>20</sup> The study of perception is arguably even more established than the study of emotion, and a distinct field has developed that is dedicated to it. People that work in this field attempt, among other things, to characterise various different kinds of perceptual state and their features.<sup>21</sup> Similarly, the study of intention has developed a degree of autonomy from Philosophy of Mind and is much discussed in Philosophy of Action, and in Practical Reasoning:<sup>22</sup>

---

see e.g. Bayne and Pacherie (2005), Coltheart (2005), Fine et al (2005), Davies (2001), Hohwy and Rosenberg (2005), Stone and Young (1997), and so on.

<sup>20</sup> For example, for cognitive theories of the emotions see e.g. Nash (1989), Neu (2000), Scheffler (1991), for perceptual theories see e.g. Clarke (1986), Griffiths (1997), (2003), James (1884), Kerner (1970), Prinz (2004), (2005), Roberts (1992), (1996), Tappolet (2005), for evaluative theories see e.g. Blackburn (1998), Greenspan (1988), Helm (2002), Hookway (2003), Oddie (2005), Nussbaum (2004), for work on the connections between emotions and rationality see e.g. Audi (1977), Badcock (2004), Ben-Ze'ev (1997), (2003), De Sousa (1987), (2003), Jones (2004), Goldie (2004), Greenspan (1988), (2003), Helm (1994), Solomon (1977), etc, for work on the relations between emotions and action see e.g. Lyons (1978), Pacherie (2002), Prinz (2004), etc.

<sup>21</sup> For example, for work on the content of perception see e.g. Ayers (2002), Bermudez (2000), (2003), Brewer (1999), Byrne (2003), Chalmers (2006), Crane (1992), Davies (1992), Kelly (2001), MacPherson (2003), (2005), Martin (2002), McDowell (1994), Millikan (1991), (1993), Pacherie (2000), Peacocke (1983), (1992), (1998), Siegel (2004), (2006), Tye (1994), (2000), for work on the connection between perception and belief see e.g. Austin (1964), Clark (1981), Evans, (1982), McDowell (1994), Millar (1991), Noe (1999), Pendlebury (1990), Soltis (1966), Tye (1995) for work on the connection between perception and action see e.g. Clark (2002), Hurley (1998), Noe (2005), for discussion of perceptual illusions see e.g. Armstrong (1968), Crane (1988), (2002), Fisher (1978), Johnston (2004), Maynard (1994), Mellor (1988), etc, for work about Direct and Indirect Realism in perception see e.g. Armstrong (1961), Austin (1964), Ayer (1940), Brewer (2004), Crane (1992), McDowell (1994), Russell (1912), (1921), Smith (1999), Thau (2004), Valberg (1992), for work on the phenomenal character of perception see e.g. Chalmers (2006), Hellie (2005), Jackson (1982), Tye (2002), etc, for work on the relations between perception and aesthetics see e.g. Crick (1976), Gombrich (1960), Gombrich, Hochberg and Black (1972), Lycan (1971), Morton (1974), Wolheim (1973), Ziff (1951),

<sup>22</sup> Significant writers here include e.g. Anscombe (1957), Audi (1973), (1991), (1993), Bratman (1984), (1987), (1991), (1993), (1999), Blackburn (1998), Broome (2004), Dancy (1999), Davidson (1980), Dretske (1988), (1999) Hampshire (1965), Harman (1983), Kavka (1983), Mele (1992), (1999), (2005), Millar (2004), Millikan (1993), Quinn (1989), Pacherie (2000), Raz (1999), (2001), Smith (1994), Searle (1979), (1983), Velleman (1989), (2000), etc.

In contrast there are few well-worked out theories about desire.<sup>23</sup> And there has been little similar research into analogous questions and problems about desire, for example, what desire has in common with other conative states like hopes and wishes and in what ways they differ, what the possible objects of desire are and how they are related to our beliefs about those things, the role of desire in motivating and actually bringing someone to act in different ways, whether there are standards of correctness distinctly applying to desire such that it can be reasonable to criticize someone for having certain desires in a similar way that someone can be criticised for having false beliefs or for believing things that are inconsistent, and so on. The importance of desires in our mental lives is not reflected in the extent to which they are studied and in the amount that is understood about them. This dissertation can contribute to setting that right.

### **Structure of the Thesis.**

The approach of this thesis is mainly defensive. I take there to be six principle kinds of objection to MN and in the following chapters I will address each in turn in detail and attempt to defend MN against it. However, I begin, in the

---

<sup>23</sup> See again footnote (15) above.



first chapter **Motivations for Motivational Necessity**, by giving some *prima facie* reasons to accept MN. First, it is presupposed by a number of important debates in various fields of philosophy and by the majority of people currently working in philosophy of mind. This alone might suggest the intuitive plausibility of the view: there would be extensive philosophical consequences if in fact MN was false. Second, MN provides a number of theoretical advantages for work within philosophy of mind in particular. I highlight some of these that have been offered by people who have explicitly accepted MN as part of a particular analysis: they appeal to MN's potential for answering some persistent problems in philosophy of mind. In addition I highlight some of the explanatory advantages that MN has with respect to certain facts about desire in particular. I argue that MN provides explanations of these facts that are better than the explanations that could be provided in terms of other features that are commonly proposed as necessary conditions for having a desire.

In the second chapter, **Motivation, Desire and the Weather Watchers**, I address Galen Strawson's objection to MN based on his example of the Weather Watchers (1994). These are creatures that, according to Strawson can have desires but cannot act or be disposed to act. He claims that it is intuitive that creatures like the Weather Watchers are possible. If Strawson is right then they are a counter example to MN: they show that it is possible for someone to have a desire and not have a disposition to act.

However, I argue that the example fails. It is under-described and although it appears to be a case in which the Weather Watchers have desires and are not disposed to act, if the additional details necessary to properly understand the example are filled in then this is not the case. I argue that there are two ways in which this can be done but that according to the first of these the Weather Watchers cannot have dispositions to act but cannot have desires either, while according to the second the Weather Watchers can have desires but if they do then they will have dispositions to act as MN claims. So neither of the ways in which the example can be properly understood is a case in which the Weather Watchers both have a desire and do not have a disposition to act: they are not counter examples to MN. However, the counter example fails for interesting reasons: it suggests that there are further relations both between desires and intentions to act and between intentions to act and dispositions to act.

Next, in **Troublesome Desires**, I address three more putative counter examples to MN. Each of these is a case in which the specific content of a particular kind of desire, its propositional object, is claimed to make it a case in which a subject who has such a desire cannot be disposed to act to try to satisfy it. The first, what I call self-passive desires, are desires to have something without having to do anything yourself to get it: for example a desire that ones child passes an exam on their own. I argue that self-passive desires are not a counter example to MN if they are properly understood:

moreover MN can accommodate these cases in a very plausible way when they are described in full detail. The second kinds of case are desires for something that is necessarily true and obviously so, like a desire that  $2 + 2 = 4$ , while the third kinds of case are desires for something that is obviously impossible, like a desire that I have a square circle. I argue that these two cases are merely specific instances of a more general problem, what Stalnaker calls ‘the problem of equivalence’ about how it is possible to represent one and not another of a set of propositions that are necessarily equivalent to one another. This is not specifically a problem for giving a theory of desire and something that we should expect to be answered in this context. Rather, it is plausible to think there must be some solution to this general problem and that if this general solution is applied to these two cases then it will show that they are not counter examples to MN.

The fourth chapter, **Defending Dispositional Direction of Fit**, is concerned with the difference between beliefs and desires. A popular way of characterizing what distinguishes these kinds of mental states is in terms of their having different directions-of-fit. But what does it mean to have a particular direction-of-fit? I argue that the most plausible kind of account analyses this in terms of a state’s functional role, its characteristic properties and relations. According to this kind of account it is necessary for a kind of state to have the direction-of-fit of desire that if someone is in that state then

they are disposed to act in ways they believe will bring about what they desire. So this plausible way of distinguishing between beliefs and desires entails that MN is true.

In the chapter I defend this account of direction-of-fit against different objections that argue, for different reasons, that it is uninformative. They argue that it fails to provide information about the features of each kind of state that are essential for its being a state of one kind and not the other, that it merely tells us about the features that are not relevant to this particular difference. But these objections fail: in one case the objection is based on a misinterpretation or failure to recognise certain significant details of the account that are relevant to this particular difference. In the other case the objection is based on implausible and overly demanding assumptions about what is required for an analysis to be informative. So these objections fail, and we should accept this way of characterizing the difference between beliefs and desires, which entails that MN is true

Finally, in **Desire and Reward** I address an objection to MN that claims that there is a conflict between MN and the kind of empirical information that we have from current neuroscience. According to Timothy Schroeder, who makes this objection, this kind of information must be incorporated into an analysis of desire if that analysis is to be informative. Yet, he claims that it supports what he calls a ‘Reward Theory of Desire’, and

according to such a reward theory MN is false. However, I argue that both of Schroeder's conclusions here are false. First, I argue that to the extent that the neuroscientific evidence supports a reward theory it also supports an analysis that entails MN is true, what we might call a 'Motivational Theory of Desire'. In addition, I argue that the most plausible versions of reward theory are, in fact also versions of motivational theory: plausible versions of reward theory themselves entail that MN is true. So MN is consistent with and supported by the empirical evidence.

I will then briefly conclude by summarizing the arguments that I have given in the different chapters of the dissertation and what I take them to show. I will also briefly discuss some of the further issues, and areas of interest for research that come up during the dissertation but that I am not able to talk about in detail.

# **CHAPTER 1:**

## **MOTIVATION FOR MOTIVATIONAL NECESSITY.**

### **1.1 – Introduction.**

What turns on the truth of MN? In the following chapters I will defend the thesis against what I take to be the main objections to it. But what is the significance of this project? In this chapter I will attempt to highlight the importance and widespread acceptance of MN. It is presupposed by a number of flourishing, keenly contested debates in different fields of philosophy, the landscape of which would be significantly altered if MN were false. In addition MN has theoretical and explanatory benefits for work in Philosophy of Mind: it provides a plausible way of answering some traditional problems within this field. And it provides the most plausible explanations of certain prominent facts about desire. So this is not a trivial issue: it matters, both for Philosophy of Mind and more broadly in philosophy, whether MN is true.

## 1.2 – MN in Different Debates.

Recall from the Introduction, I described a range of fields of philosophy that appealed to the notion of desire in different debates. I described debates between Internalist and Externalist views in both Meta-Ethics and in Practical Reasoning, and a debate in Aesthetics and Philosophy of Language between different versions of make-believe, or pretense accounts of fiction and our reactions to works of fiction. In fact, these debates not only appeal to the notion of desire but they presuppose the truth of MN.

Consider again the example from Meta-Ethics: this is where the debate between Internalists and Externalists and the debate between Descriptivists and Expressivists intersect. It is often considered the crucial issue within Meta-Ethics and is what Smith calls ‘The Moral Problem’ (Smith 1994). He presents the issue as the apparent conflict between three common and powerful intuitions about moral judgments and about human psychology: first, what Smith calls ‘the objectivity of moral judgment’ which suggests that

there are moral facts, wholly determined by circumstances, and that our moral judgments express our beliefs about what those facts are (Smith 1994, p11).

Second, 'the practicality of moral judgment' which suggests that

moral judgments seem to be, or imply, opinions about the reasons we have for behaving in certain ways, and, other things being equal, having such opinions is a matter of finding ourselves with a corresponding motivation to act (Smith 1994, p7).

Finally, according to 'the standard picture of human psychology' [due to Hume (1975)]

there are two main kinds of psychological state. On the one hand there are beliefs, states that purport to represent the way the world is. Since our beliefs purport to represent the world, they are assessable in terms of truth and falsehood, depending on whether or not they succeed in representing the world to be the way it really is. And on the other hand there are desires, states that represent how the world is to be. Desires are unlike beliefs in that they do not even purport to represent the way the world is. They are therefore not assessable in terms of truth and falsehood. Hume concludes that belief and desire are therefore distinct existences: that is, that we can always pull beliefs and desire apart, at least modally. For any belief and desire pair that we imagine, we can



always imagine someone having the desire but lacking the belief, and vice versa (Smith 1994, p7).

In addition, this standard picture provides the following model for explaining human action:

Crudely our beliefs tell us how the world is, and thus how it has to be changed, so as to make it the way our desires tell us it is to be. An action is thus the product of these two distinct existences: a desire representing the way the world is to be and a belief telling us how the world has to be changed so as to make it that way (Smith 1994, p9).

The idea is that beliefs cannot motivate someone to act, rather that only desires can do this, so to be motivated to act in a particular way it is necessary that someone has both a desire and a belief that that desire can be satisfied by acting in that particular way.

The Moral Problem then is that it seems as though moral judgments express beliefs about facts and it also seems that having such a belief either itself motivates someone to act or entails that they have a corresponding desire that provides this motivation when had along with such beliefs. Yet we think that beliefs cannot do either of these things: they neither motivate people to act

nor are they necessarily connected with having particular desires. This conflict suggests that, in fact, we cannot make moral judgments as there is no psychological state that we can have that has the properties necessary to be such a judgment. Hence, the Moral Problem threatens us with moral nihilism if it cannot be solved (Smith 1994, p13).

A range of responses have been made to try to solve the Moral Problem. Some like Frankena (1958), Foot (1972), Scanlon (1982), Railton (1986), Brink (1989), Copp (1997), Svavarsdottir (1999) and Shafer-Landau (2003) respond by denying the practicality of moral judgment: such externalists deny that making a moral judgment necessarily motivates someone to act. Rather they claim that the connection with being motivated to act is 'external' to making a moral judgment, that whether or not someone will be motivated depends on what other mental states, like desires, they have.

Others like Ayer (1936), Carnap (1937), Hare (1952), Mackie (1977), Blackburn (1984), and Gibbard (1990), respond by denying the objectivity of moral judgment: such expressivists deny that making a moral judgment is a matter of forming a belief about certain facts. They claim that moral judgments are actually expressions of, for example, our approval or disapproval of certain things and ways of behaving, or expressions of our desires for such things, or commands to others to behave in certain ways, and so on. They typically claim that it is because the surface grammar of moral

judgments is like that of expression of beliefs that we are misled into thinking that moral judgments express our beliefs.

Others still, like Nagel (1970), McDowell (1978), Platts (1979), McNaughton (1988), Darwall (1983), Dancy (1993), Scanlon (1998) and Shafer-Landau (2003) reject the standard picture of human psychology: these 'Anti-Humeans' typically claim that having a belief either can be sufficient for being motivated to act or that it can entail having a particular desire. So according to anti-Humeans making a moral judgment can be an expression of one's beliefs, and it can motivate someone to act in accordance with their judgment.

I do not intend to enter into this debate here. Rather what is illustrative for my purposes is that all sides of the debates typically assume that if those positions that claim that making a moral judgment is a matter of having a desire to act in accordance with that judgment, or if it entails having such a desire then the practicality of moral judgment follows straightforwardly from this. It is a presumption of the debates that if someone has a desire to act in accordance with a moral judgment then necessarily they are motivated to act in that way. For instance, even externalists typically accept that if expressivism is true and moral judgments express our desires then the practicality of moral judgment follows from this, and internalism is true. But they will deny that such versions of expressivism are true. On the other hand,

even expressivists typically accept that if externalism is true and moral judgments are not necessarily connected with being motivated to act then expressivism, or at least those versions that claim that moral judgments are expressions of desires, is false. But they will deny that externalism is true. And even Anti-Humeans typically accept that if expressivism is true then their denial of the standard picture of human psychology is unmotivated. That is, they accept that if moral judgments are expressions of desire then this is sufficient for there to be a necessary connection between making a moral judgment and being motivated to act. There would be no need, in this context at least, to deny the standard picture. So, all sides of this debate typically presuppose that there is a necessary connection between having a desire and being motivated to act. If we understand someone being motivated to act in terms of their having a disposition to act in ways that they believe will bring about what they desire [as for example Smith does (1994)] then it is MN that is presupposed by this debate.

The analogous debate between in Practical Reasoning presupposes MN in a similar way. Indeed, moral judgments are often considered to be just a particular example of judgments about what reasons you have.<sup>24</sup> Analogous

---

<sup>24</sup> See for example Rosati (2006) who says, 'Moral motivation is an instance of a more general phenomenon—what we might call *normative motivation*—for our other normative judgments also typically have some motivating force. When we make the normative judgment that something is *good for us*, or that we have a *reason* to act in a particular way, or that a specific course of action is the *rational* course, we also tend to be moved.' (Rosati, C.S. 2006: 'Moral Motivation': Stanford Encyclopedia of Philosophy, accessed online

positions are available with respect to judgments about your reasons to act in particular ways as are available with respect to moral judgments, with analogous relations between the different views. For example, Internalists about practical reason judgments typically claim that judging that you have a reason to act in a particular way is necessarily connected with being motivated to act in that way. In contrast, Externalists about practical reason judgments deny this and typically claim that whether or not you will be motivated to act in accordance with your judgments about your reasons depends on whether you have, in addition, desires to act in this way, or that you believe will be brought about by acting in this way. This debate intersects with the debate between Expressivists about practical reason judgments, who will say these judgments are an expression of approval of acting in that way, or of a desire to act in that way, for example and Descriptivists who say that these judgments are expressions of beliefs about what reasons you have. Again, these debates presuppose MN in an analogous way to the debates in Meta-Ethics.

Finally, consider the debate within Aesthetics and Philosophy of Language between different make-believe or pretense accounts about fictions and our responses to fictions. Recall, the relevant debate is between those views that attempt to explain our responses to fiction in terms of our having genuine beliefs and desires, for example beliefs about how people tend to act

---

<http://plato.stanford.edu/entries/moral-motivation/> (17-02-2007)). See also (Smith 1994, pp. 61-62.)

in the fictional situations and desires to act in similar ways, and those views that attempt to explain these responses in terms of our having merely belief-like and desire-like attitudes about and towards the fictions. One of the arguments given by those who propose the second kind of account appeals to the fact that people are rarely motivated to act in response to fictions in the same ways that they would be motivated to act in similar real situations. For example, someone who is watching *Invasion of the Body Snatchers* is not motivated to actually run away, or to actually call the police, and so on. The presumption seems to be that if someone had the kind of genuine desires when responding to fictions that the first kinds of account propose then this would entail that they had such motivations. Hence, they claim that what people have instead are attitudes that are desire-like: and one of the ways that these desire-like attitudes are said to differ from genuine desires is in terms of lacking a necessary connection with being motivated to act. In contrast, those who give the first kind of account attempt to explain this apparent lack of motivation away in terms of the content of their beliefs about the appropriate responses to fictions, or in terms of their having other desires that entail motivations to act in ways that conflict with these and over-ride their dispositions to act (say, desires to not be embarrassed or desires to act like the others in the cinema, and so on). So again, it is presupposed in this debate that

having a desire entails being motivated to act, and if this is understood in terms of having a disposition to act then it is MN that is presupposed.

As we can see, it is not merely desires that have widespread importance in philosophy but MN itself has similar widespread importance. These lively and important debates appear to presuppose that MN is true, they presuppose that there is a necessary connection between having a desire and being disposed to act in ways you believe will bring about what you desire. So people involved in such debates should also be interested in defending MN. If it turned out to be false it is likely that there would be significant implications for those debates, with respect to the relations between and plausibility of different views, the kinds of argument that are or might be relevant to them, and so on.

### **1.3 – MN in Philosophy of Mind.**

As I have described, MN is presupposed by those working in a number of fields of philosophy. In a similar way the common consensus among those working in Philosophy of Mind is that MN is true.<sup>25</sup> Indeed, it is sometimes

---

<sup>25</sup> For example, Schroeder calls the theory that having a desire consists in having a disposition to act ‘the Standard Theory’. According to this theory it is both necessary and sufficient for having a desire that someone is disposed to act in ways they believe will bring about what

taken to follow from adopting a functionalist approach to analysing mental states, which is currently the most popular and plausible approach. Of course, MN is not entailed by the functionalist approach itself: functionalism, in philosophy of mind for example, is just a technique for deriving necessary and sufficient conditions for having a particular kind of mental state from a collection of claims about the features and relations of that kind of state. So what is entailed by a functionalist account of a particular kind of mental state depends on the initial claims that the account is derived from. MN will be entailed by a particular functionalist analysis only if it was presupposed beforehand.

To see this, consider how someone might proceed to give a functional analysis of desire.<sup>26</sup> Suppose they begin by distinguishing the following two claims that they take to describe the essential features and relationships of desire: that people are motivated to act to try to satisfy their desires, and that people are usually pleased if their desires are satisfied. The first step towards an analysis of desire is to make these claims more precise. Suppose this yields the following:

---

they desire, so it entails MN. He attributes the lack of recent research into desire on the fact that this theory is so established (Schroeder 2004, pp. 10-11).

<sup>26</sup> Thanks to Daniel Stoljar: this way of presenting the functionalist technique here is directly due to discussions with him.



If  $S$  desires that  $p$  then  $S$  is disposed to act in a way  $S$  believes will bring about  $p$ .

If  $S$  desires that  $p$  then  $S$  is pleased if  $S$  believes that  $p$ .

Next they conjoin these claims and rearrange this conjunction around ‘ $S$  desires that  $p$ ’:

If  $S$  desires that  $p$  then  $S$  is disposed to act in a way  $S$  believes will bring about  $p$  and  $S$  is pleased if  $S$  believes that  $p$ .

The third step is to identify all of the psychological terms in this conjunction and replace them with appropriate variables to yield the following open sentence:

If  $S$   $R1$ ’s that  $p$  then ( $S$  is disposed to  $R2$  in a way  $S$   $R3$ ’s will bring about  $p$  and  $S$  has  $F$  if  $S$   $R3$ ’s that  $p$ ).

The penultimate step is to generalize this by existentially quantifying:

$\exists(R1, R2, R3, F)$  If  $S$   $R1$ ’s that  $p$  then ( $S$  is disposed to  $R2$  in a way  $S$   $R3$ ’s will bring about  $p$  and  $S$  has  $F$  if  $S$   $R3$ ’s that  $p$ ).

Finally, this is presented in the form of necessary and sufficient conditions for having a desire:

*S* desires that *p* iff:

$[\exists(R1, R2, R3, F) \text{ If } S \text{ } R1\text{'s that } p \text{ then } (S \text{ is disposed to } R2 \text{ in a way } S \text{ } R3\text{'s will bring about } p \text{ and } S \text{ has } F \text{ if } S \text{ } R3\text{'s that } p)] \text{ and } [S \text{ } R1\text{'s that } p]$ .

This is sometimes called the Ramsey sentence for this analysis. It effect it says that having a desire consists in being in a state that has a particular set of properties and relations. Which state this actually is in a particular kind of individual, or in the same individual at different times, is a further issue about what realizes desire in that kind of individual. But it should be obvious that a particular functionalist analysis of desire will only entail MN if something like the first claim I used here is among the claims that the analysis begins from. Presumably if someone denies MN then they will deny that the correct analysis of desire incorporates this claim.

Nonetheless, MN is widely accepted in Philosophy of Mind. Among the reasons for this, discussed by Smith (1994) and Stalnaker (1984) for example is that MN has significant theoretical benefits for work in the field: it

can provide a way of giving an analysis of two traditionally problematic features of our mental lives that is consistent with naturalism.<sup>27</sup> Naturalism is the view that all the facts about the world are natural facts: that is, the view that everything that exists can potentially be explained in the terms of the natural sciences, like physics, chemistry and biology, or the social sciences, including psychology (Smith 1994, p17).<sup>28</sup> There is an overwhelming consensus that naturalism is true, so any adequate theory of the mind, and of desire, must respond to these problems.

What are these traditional problems that Smith and Stalnaker highlight? They are both examples of what is sometimes called ‘the Problem of Intentionality’ or ‘Brentano’s Problem’ after the Italian philosopher Brentano who discussed it in the 19<sup>th</sup> Century (Field 1978, p9). Stalnaker describes the problem in the following way:

For various familiar reasons, intentional or representational relations seem unlike the relations holding between things and events in the

---

<sup>27</sup> Note that it is not being claimed that these two are the only features commonly associated with desire that have traditionally proven difficult to explain in a way that is consistent with naturalism. Also, these are not the only reasons that have been offered as motivation for accepting MND. Smith for instance also highlights that MND can explain a number of other features of desires, such as the fallibility of our epistemic access to our own desires and the possibility of particular desires both having or lacking an associated phenomenal character [see Smith (1994) pp. 108-115 for discussion]. These particular reasons are presented here as examples of the kind of motivations that particular theorists have offered in support of MND.

<sup>28</sup> See also Moore (1903), Ayer (1936), Wiggins (1987), etc.

natural world: causal interactions, spatiotemporal relations, various notions of similarity and difference. One can, it seems, picture, describe, or think about such things as gods and golden mountains even if they do not exist. And one can picture, describe, or think about a triangle or a sunset without there being any particular triangle or sunset that is pictured, described, or thought about. Some philosophers have used these distinctive features of intentional relations to argue that they are irreducible to natural relations. From this conclusion it is argued that mental phenomena cannot be a species of natural phenomena. Any account of thinking things as natural objects in the material world, these philosophers argue, is bound to leave something out. The challenge presented to the philosopher who wants to regard human beings and mental phenomena as part of the natural order is to explain intentional relations in naturalistic terms (Stalnaker 1984, p6).

Brentano himself denied that this challenge could be met. He claimed that necessarily mental states have intentional relations<sup>29</sup> but that intentional relations are semantic relations, and that these cannot be reduced to natural relations. Hence, Brentano thought that the mind is not purely physical and that naturalism, or materialism is false. However, the current view is that

---

<sup>29</sup> This is expressed in the slogan often taken to be Brentano's main idea, that intentionality is 'the mark of the mental'.

Brentano was wrong to conclude this. The overwhelming consensus is expressed by Field when he says

I take it as unquestionable (given what we know about the world) that materialism is true. I also take it as unquestioned [...] that people do believe and desire. These two assumptions together amount to the assumption that Brentano's problem can be solved (Field 1978, p9).<sup>30</sup>

We see a particular instance of this problem in philosophy of mind in the problem of characterising the relationship between mental states like belief and desire and their objects. As I discussed in the Introduction, it is common to describe mental states like belief and desire as having an object: when someone has a particular belief there is something that they believe is the case. And when someone has a particular desire there is something that they desire to have, or desire to be the case. But such objects cannot be concrete states of the world for those familiar reasons mentioned by Stalnaker above [and in more detail in Stalnaker (1984), and by Field (1978)] and that I discussed

---

<sup>30</sup> Two caveats deserve mention here: first although there is in fact some disagreement about the possibility of giving an analysis of certain kinds of mental states in material terms, in particular those mental states with phenomenal character like being in pain or having a visual sensation of red, it is almost uncontroversial that propositional attitudes like beliefs and desires can be analysed in material terms. Second, although some people [see for example Churchland (1981)] do deny that people have beliefs and desires, such Eliminativist theories of belief and desire are very much a minority view and are generally taken to be implausible [see for example Braddon-Mitchell and Jackson (1996) pp. 242-247 for discussion].

briefly in the Introduction. These reasons suggest that the objects of beliefs and desires must be intentional objects, or representations of some sort. The relations between a person and the objects of their beliefs and desires are therefore intentional relations. Someone trying to give an account of belief or desire therefore faces Brentano's problem if they are trying to characterise these relations.

A common response here, and one I am adopting in this dissertation, is to say that the objects of beliefs and desires are propositions: the problem then becomes the problem of characterizing the relation between a person and a particular proposition in terms of a natural relation. But merely saying that the objects of belief and desire are propositions does not suffice to provide any obvious solution. An answer to Brentano's problem must give an account, in naturalistic terms of the relational property that someone has when they have a belief that  $p$ , a desire that  $p$ , or some other kind of intentional mental state, where  $p$  stands for the proposition that is the object of that state. The standard way of characterizing a proposition is as a set of possible worlds at which the proposition is the case. But to say that the object of someone's belief or desire is the set of possible worlds where a particular proposition is true is merely to say what has this relational property. Saying this does not show that this relation can be analysed in terms of a natural relation, and nor does it say what

this natural relation is like. So it does not provide an answer to Brentano's Problem (Field 1978, p11).

The relevance of MN here is that it does provide a way of answering this problem, at least for the case of characterizing the relation between a person and the object of their desire. According to MN if someone desires that  $p$  then they are disposed to act in particular ways in particular conditions. What this is often taken to mean is that they have a certain intrinsic property that will cause them to act in such ways if the appropriate conditions obtain. For example, such intrinsic properties might be certain physical states, typically caused in particular ways and that typically cause the person to be in certain other physical states, some of which cause them to act in particular ways in certain circumstances. Although MN need not be interpreted in this way, if it is then it says that if someone has a desire that  $p$  then they have a particular set of physical and causal properties and relations. Furthermore, this is a different from the set of physical and causal properties and relations they have if that have a desire that  $q$ . But physical and causal properties are uncontroversial examples of natural properties and relations. So MN offers a way of giving a naturalistic analysis of desires having a particular object. As we can see, in the context of this problem of analysing intentional relations

this is a significant theoretical benefit and it is often given as motivation for accepting MN.<sup>31</sup>

The second of these traditional problems that I will mention here is similar, again a particular example of Brentano's Problem: it is the problem of giving a characterisation of being motivated to do something. This problem is particularly relevant to the project of giving a characterisation of desires as desires have traditionally been said to have a relation with motivation (recall, for example the passages from Hume and Locke in the Introduction). One natural way of understanding what it is for someone to be motivated in one way or another is in terms of their having a goal. Indeed, according to Smith, to be motivated at a particular time just is to have a goal at that time (Smith 1994, p116). But this is problematic as having a goal is a teleological notion: it involves being directed towards something, or having a purpose. But analogous problems arise here as discussed above concerning what kind of thing the goal of motivation can be, what it is that someone is directed towards when they have a purpose: for example it is possible that the goal does not exist at the time someone is motivated towards it, and so on. So there is a similar problem of giving an analysis of motivation that is consistent with naturalism. MN again offers a way of giving such a naturalistic analysis. It

---

<sup>31</sup> See for example Stalnaker (1984): he offers this reason in support of MN (or more precisely in support of a view that entails MN), when he says '[d]esires have determinate content because of their dual connection with belief and action' (Stalnaker 1984, p19).



characterizes being motivated towards a particular goal in terms of having a desire for it, that is, characterizing this goal as the object of a desire. Then in turn MN claims that if someone has a desire for this object then they are disposed to act in certain ways. And as discussed above this can be understood in terms of having a particular set of physical and causal properties and relations, which are natural properties and relations. So MN offers a way of giving a naturalistic analysis of motivation: this is another significant theoretical benefit and is similarly given as another reason to accepting the thesis.<sup>32</sup>

MN is widely accepted within Philosophy of Mind and there are good philosophical reasons for this. It provides ways of giving naturalistic analyses of certain important features discussed in the field, like the relations between people and the objects of their mental states, and the property of being motivated in a way. This is important as these features have proven persistently difficult to analyse in a naturalistic way. So there are significant theoretical benefits for general work in Philosophy of Mind if MN is true

#### **1.4 – A Fact About Desire.**

---

<sup>32</sup> See for example (Smith 1994, p103 and p115).

In addition to these benefits for Philosophy of Mind in general, if MN is true it has explanatory advantages that are particularly relevant when attempting to give an analysis of desire: MN provides the most straightforward explanation of a certain, very prominent fact about desire. It is not obvious that this can be explained in terms of the properties and relations that are said to characterise desire by standard, alternative theories of desire that reject MN.

It is a fact that having a desire is regularly correlated with acting in ways that you believe will bring about what you desire. This is empirically established and widely acknowledged: it is plausible that our recognition of this correlation underlies our everyday practices of trying to explain peoples' behaviour by attributing them with certain desires, and making predictions about peoples' behaviour on the basis of what desires we believe they have. As Stalnaker describes,

[i]t is, I think, intuitively clear that however often we may fail to act according to our beliefs [about ways that we could act that would satisfy our desires<sup>33</sup>], there is a presumption that we do. Where people don't do what is appropriate, given their beliefs, we expect there to be some explanation for this; we may appeal, for example, to incapacity, absentmindedness, or self-deception (Stalnaker 1984, p19).

---

<sup>33</sup> See, for example (Stalnaker 1984, p15) for his account of beliefs as dispositions to act in ways that would tend to satisfy one's desires if one's beliefs are true.

It is also plausible that the widespread recognition of this correlation can explain why it is that we take acting in certain ways to constitute evidence that someone has particular desires. As Anscombe says, '[t]he primitive sign of wanting is *trying to get*' (Anscombe 1957, p68).<sup>34</sup> Anscombe here uses 'want' as roughly synonymous with 'desire'. She seems to be saying that we consider the ways that someone acts, along with what we take them to believe about what their actions will achieve, to be the most basic evidence of what their particular desires are. This seems uncontroversial: not only do we pay particular attention to the way that people actually act when attributing desires but also how they act is sometimes taken as more reliable evidence of their desires than what they say about what it is they desire. For example, even if someone says that they desire to exercise more and get fit we will tend to doubt that they do really desire this if they take no steps to get more exercise even when given a number of opportunities and where there are no obvious

---

<sup>34</sup> It seems that Anscombe's aim in the passage that this is taken from is to distinguish desire as the pro-attitude that is relevant to practical reason and action, as opposed to other pro-attitudes like hopes or wishes that she says are not connected in this way with acting. However, it is the claim about someone's actions being 'the primitive sign of' desire that is relevant here: she seems to be saying that how someone acts is the most basic evidence of their having certain desires. Indeed, Anscombe may even be claiming that there is a stronger connection between desiring and try to get than just an evidential connection. A different way of understanding the claim that *a* is a *sign* of *b*, is as a claim that *b* is partly constituted by *a*, that *a* is necessary for *b*. It is in this way that the term 'sign' is sometimes used in medicine, for example where a distinction is drawn between signs of a certain disease or condition, which are diagnostic criteria for it, and symptoms of a certain disease or condition, which are evidence for it. However, I will here take Anscombe to be making only the weaker claim, that there is an evidential connection between desiring and trying to

factors that might interfere with their doing so (such as the lack of affordable and convenient facilities for taking exercise, pressing deadlines at work, commitments to other projects, and so on). Moreover, we think that these doubts are legitimate, that the behavioural evidence can outweigh the testimonial evidence in such cases. Of course, we cannot straightforwardly conclude from this that it is their claims about what they desire that are false when there is this conflict between different kinds of evidence: for example, the mismatch might be explained in terms of their having mistaken beliefs about how to act to satisfy the desires that they claim to have. And even if we do reach the straightforward conclusion this does not show that we are being insincere in making those claims as there are other possible explanations: for example, they may be self-deceiving, or may be ignorant of what they really want. But consistent failure to act to try to satisfy a desire when there are no obvious factors preventing someone from doing so is taken as good evidence that they do not actually have such a desire.

It is plausible that a person's actions are taken to be such good evidence for what they desire because there is a regular correlation between having a desire for something and acting to try to satisfy that. Of course, this correlation is not exception-less: there might be a number of reasons why someone who has a particular desire does not actually act in a particular way

---

get. The stronger claim is, according to a natural interpretation of trying to get in terms of being disposed to act, just the thesis MN.

as a result of having that desire. For example, they might lack any opportunities to act on it or fail to recognize such an opportunity, they might have other, stronger desires that they act on instead, and so on. But it is uncontroversial that there is such a regular correlation, and this is a fact that stands in need of explanation.

### **1.5 – Explaining this Fact.**

This regular correlation between having a desire and acting in ways you believe will bring about what you desire is among the prominent phenomena associated with having a desire. So it is something that should be explained by an analysis of desire: it is reasonable to expect that a specification of the necessary and sufficient conditions for having a desire will explain why desires are associated with certain other phenomena in particular. This particular fact about desire, its correlation with action, is better explained in terms of MN than it is explained in terms of the other features that are standardly proposed as necessary conditions for having a desire. This is good reason to accept that MN is true.

What different properties and relations do the standard competing theories of desire propose for giving their respective analyses? And how

might they each explain this fact? Recall, MN is not a theory of desire: it is not a set of necessary and sufficient conditions for having a desire. Rather it is just a necessary condition. There do not seem to be any prior restrictions on what other necessary conditions a particular theory can claim are sufficient for having a desire in addition to MN: so there are no prior restrictions on how different theories of desire that have in common that they entail MN can differ from one another in terms of their other necessary conditions. Call any theory that entails MN a Motivational Theory of Desire: all motivational theories entail that having a desire entails having a disposition to act in ways you believe will bring about what you desire.

Among non-motivational theories, those that do not entail MN, what different properties and relations are claimed as being necessary conditions for having a desire? One traditional view is that there is a necessary connection between having desires and having experiences of pleasure. This is proposed by what are sometimes called Hedonic Theories of Desire, the kind of theory that has been given by, for example Mill (1863), Kim (1998), and Fehige (2001) for example. As with motivational theories, no plausible hedonic theory will claim that having a desire entails having an experience of pleasure: rather they will say that it entails being disposed to have an experience of pleasure when they have certain thoughts or beliefs about the object of their desire. Different versions of hedonic theory have in common that they entail

something like the following claim: necessarily if  $S$  desires that  $p$  then  $S$  is disposed to have an experience of pleasure if  $S$  believes that  $p$  has come about.

A different kind of account claims that there is a necessary connection between having a desire for something and making an evaluation of that object. Different versions of this kind of account, that might be called Evaluative Theories of Desire, vary with respect to what they say the nature of this evaluation is: for example, some versions say that having a desire entails having an evaluative belief about what you desire,<sup>35</sup> while others say that having a desire entails having a non-cognitive evaluative state of some kind,<sup>36</sup> often said to be like an evaluative perception of the object of desire. Different versions also vary with respect to what evaluation is made of the object of desire: for example, some versions say that having a desire entails evaluating the object of desire as good, others say that having a desire entails evaluating it as valuable, or desirable, and so on. But different versions of evaluative theory tend to agree that the object of desire is evaluated in a positive way.<sup>37</sup>

---

<sup>35</sup> See for example Lewis (1988): according to what Lewis calls the Desire as Belief (DAB) theory  $S$  desires that  $p$  iff  $S$  believes that  $p$  is good.

<sup>36</sup> See for example Oddie (2005) who says that having a desire that  $p$  is constituted by having an experience of  $p$  as being valuable (where this is analogous to perceptual experiences)

<sup>37</sup> It is plausible that such theories are motivated by discussions by people like Anscombe, for example, about the possible objects of desire: Anscombe famously claimed that it always makes sense to ask about the object of a desire what its desirable characteristics are, which is often taken to suggest that it is not possible to desire something unless you take it to have some positive characteristic (either intrinsically or in virtue of being a means to getting something that you take to have positive characteristics intrinsically) (Anscombe 1957, p70).

So we can characterise evaluative theories as having in common the claim that if  $S$  desires that  $p$  then  $S$  represents  $p$  in a positive evaluative way.

Another, more recent proposal, by people like Dretske (1988) and Schroeder (2004), attempts to analyse desire in terms of relations with rewards. These Reward Theories of Desire typically claim in common that if  $S$  desires that  $p$  then  $S$  represents  $p$  as a reward. But versions of this kind of theory may differ with respect to what it is to represent something as a reward: Schroeder for example, characterizes this in terms of the notion of reward commonly used in behavioural science, that is, something is a reward for someone if it has characteristic effects on them by processes of conditioning, or reinforcement learning.

As far as I am aware, these four kinds of theories are the standard alternatives for analysing desire: particular theories tend to be a version of one or another (or more than one) of these theories. They attempt to analyse desire in terms of relations with dispositions to act, experiences of pleasure, evaluations or rewards. The other proposals about how to characterise desire that are not theories of one of these kinds are more accurately described as denying that it is possible to analyse desire. For example, Galen Strawson claims that having a desire is not necessarily connected with any other kind of property or relation. He claims that any such connections that desires have to other phenomena are merely contingent [Strawson (1994) (1998)]. According



to Strawson what makes a mental state a desire that I have some ice cream, as opposed to a belief that I have some ice cream, is the fact that it is a desire, rather than a belief. And what makes it a desire that I have some ice cream as opposed to a desire that I have some chocolate is the fact that it is a desire for ice cream and not chocolate. We might call this a Primitivist Theory of Desire where according to primitivist theories *S* desires that *p* iff *S* has a sui generis state that is a desire that *p*.

The other kind of ‘non-analysis’ would be given by an Eliminativist Theory of Desire, as held by Churchland (1981), for example. As the name suggests eliminativists deny that there are such things as desires: they typically claim something like, there is no particular state *x* such that *S* has *x* when *S* is described as having a desire that *p*. Eliminativists such as Churchland tend to claim that the term ‘desire’ is just a term from an outdated folk-psychological theory of the mind and that it does not actually refer to a kind of psychological state that we actually have. Rather they claim that this folk-psychological theory will be superceded by advanced neuroscience, and the mental state terms used in this folk psychology, like ‘desire’, should be replaced with those of this neuroscience to give a more accurate account of the different mental states and processes that we actually have. There will be nothing in this account that fits the folk-psychological concept well enough to deserve the name ‘desire’: so in effect there are no desires.

Clearly these last two alternatives cannot explain the fact that there is a regular correlation between having a desire and acting in ways you believe will bring about what you desire. A primitivist claims that this is merely a contingent fact and that it is completely independent of the kind of mental state desire is. On the other hand an eliminativist denies that this is a fact at all: they deny that there is a particular kind of mental state, a desire, that these actions are regularly correlated with. For an eliminativist there is no fact to explain whereas for a primitivist the fact is not in need of any special explanation.

But how do the other alternative kinds of theory compare? Which of the four different relations provides the best explanation of this fact? First consider motivational theories. According to motivational theories MN is true, so they claim that if someone has a desire then they are disposed to act in ways they believe will bring about what they desire. Of course, this disposition will only manifest in their actually acting in this way in certain circumstances, if certain conditions obtain. But if we assume that such conditions can obtain quite regularly then if someone has a desire this will regularly result in their acting in a way they believe will bring about what they desire. On this assumption this fact, that there is a regular correlation between having a desire and acting in these particular ways follows straightforwardly from

motivational theories. So if MN is true there is a straightforward explanation of this fact about desire.

In contrast neither hedonic theories, evaluative theories, nor reward theories provide any explanation of this fact. Note, this is not to say that any of these kinds of theories is incompatible with there being an explanation: they are not like primitivist theories of desire that deny that there is anything interesting to say about the regular correlation between desire and action. Rather, these theories leave the correlation unexplained. This is because there is no obvious connection between acting and the properties and relations that each kind of theory claims is necessary for having a desire. Recall, hedonic theories, evaluative theories and reward theories claim respectively that having a desire that  $p$  entails being disposed to have an experience of pleasure if you believe that  $p$  has come about, or that you represent  $p$  in a positive evaluative way, or that you have representations of  $p$  as a reward (such that these representations have characteristic effects on you by processes of reinforcement learning, for example).<sup>38</sup> Yet none of these properties and relations have an obvious connection with acting. So, saying that having such a property or relation is necessary for having a desire does not say anything that is obviously relevant to the regular correlation between having a desire and acting in ways you believe will bring about what you desire. It might be

---

<sup>38</sup> Of course, there will be some variation about what specifically these conditions are between different versions of each kind of theory.

possible for theories of these kinds to explain this fact: for example, someone who held a hedonic theory might claim that there is a necessary connection between being disposed to have experiences of pleasure if you believe that something has come about and acting in ways you believe will bring it about. Analogously, an evaluative theorist might claim that there is a necessary connection between representing something in a positive evaluative way and acting in ways you believe will bring it about. And a reward theorist might claim that there is a similar connection between representing something as a reward and acting in these ways: for example, someone like Schroeder might claim that one of the characteristic effects that representations that  $p$  have on you if you desire that  $p$  is that they condition you to act in ways that you believe will bring about  $p$ . By making some kind of claim like these the people who held these different kinds of theory could thereby explain why desire is regularly correlated with particular actions. But note that to do this they must make a further claim about the property or relation that they claim is necessarily connected with desire. Note also that these further claims themselves must be argued for as none of these connections between the respective features and actions is true a priori. So none of these kinds of theory can explain the regular correlation between desire and particular actions in virtue of the property or relation that they claim is necessary for having a desire: by saying that having such-and-such property or relation is necessary

for having a desire they do not thereby explain the fact we are interested in here. This is in contrast to motivational theories of desire as we saw above: this fact follows in a straightforward way from the truth of MN. So, motivational theories of desire provide the best explanation of the regular correlation between having a desire and acting in ways you believe will bring about what you desire. As this correlation is one of the prominent phenomena associated with having a desire, something that should be explained by an analysis of desire, this is a significant benefit of motivational theories: it is good reason to think that MN is true.

## **1.6 – Conclusion.**

There is a widespread presumption both in philosophy in general and in Philosophy of Mind in particular that MN is true. It is presupposed by a number of important debates in various fields of philosophy. And there are good philosophical reasons to accept it: it provides a way of giving plausible accounts of certain features that are important in Philosophy of Mind and that have proven persistently difficult to analyse in naturalistic terms. In addition, it provides the most straightforward explanation of certain facts about desire from among the alternative proposed analyses: this is something that is

especially relevant to the issues in this dissertation. So there would be significant implications for the possibility of giving an adequate analysis of desire, for issues in Philosophy of Mind in general, and for philosophy more broadly if MN turned out to be false. All of this is good reason to think that MN is true and to be concerned to defend it against the various, seemingly persuasive, objections made against it.

# CHAPTER 2

## MOTIVATION, DESIRE AND THE WEATHER WATCHERS.

### 2.1 – Introduction.

The Weather Watchers are a race of sentient, intelligent creatures. They are distributed about the surface of their planet, rooted to the ground, profoundly interested in the local weather. They have sensations, thoughts, emotions, beliefs, desires. They possess a conception of an objective, spatial world. But they are constitutionally incapable of any sort of behavior, as this is ordinarily understood. They lack the necessary physiology. Their mental lives have no other-observable effects. They are not even disposed to behave in any way (Strawson 1994, p251).

According to Galen Strawson, who gives this example, the Weather Watchers are ‘an evidently possibility’ (Strawson 1994, p256). They are a counter example to MN: they can have desires but they cannot act and they cannot be

disposed to act yet MN says that having a desire entails having such a disposition to act. So if Strawson is right and the Weather Watchers are possible then MN is false.

However, in this chapter I argue that the Weather Watchers are not possible, contrary to Strawson's claim: the case is under-described and on closer examination the example is not one in which they both have desires and lack dispositions to act. So there is no counter example to MN. But the reason the example fails is revealing: it suggests that there are interesting relations between desires and intentions, and between intentions and dispositions to act. I will attempt to explicate these relations while defending MN against Strawson's objection.

## 2.2 – Strawson's Objection.

As we have seen, Strawson argues by counter example against MN. His objection is, on the surface, like the following simple argument:<sup>39</sup>

---

<sup>39</sup> Strawson's stated target with the example is actually a view he calls 'Neo-Behaviourism'. This is the view that 'mental life is linked to behavior in such a way that reference to behaviour enters essentially and centrally into any adequate account of the nature of almost all if not all mental states and occurrences, like emotions, sensations, thought, beliefs and desires' (Strawson 1994, p29). From this it is not clear exactly what follows from the view about desire in particular. However, in the symposium on *Mental Reality in Philosophy and Phenomenological Research* (June 1998) Michael Smith characterizes Neo-Behaviourists as claiming 'that a subject's desire that p is a state that necessarily manifests itself in her behaviour, though perhaps only indirectly' (Smith 1998, p450). Again, 'manifesting' in



- 1) If MN is true then it is not possible that *S* desires that *p* and *S* is not disposed to act in ways *S* believes will bring about *p*.
- 2) It is possible that *S* desires that *p* and *S* is not disposed to act in ways *S* believes will bring about *p*.
- C) So MN is false.

But if someone accepts MN then they will respond by denying premise 2, so why does the objection not straightforwardly beg the question against them? Strawson gives the Weather Watchers example to support this premise. But the way that it is supposed to provide this support is subtle and needs explicating in detail. This is my aim in this section and the next.

Recall from above, the features in Strawson's description of the Weather Watchers that are relevant here are the following: first that the Weather Watchers can have desires, and second, that they cannot actually perform any behaviour or have a disposition to do so. By 'behaviour, as this is ordinarily understood' Strawson means 'actions' as I characterized them in the Introduction, that is someone's intentional bodily movements as opposed to movements that merely happen to them, like reflexes. He says,

---

behaviour could be understood in a number of different ways. However, the rest of the text suggests that it should be read in terms of having a disposition to act. So Neo-Behaviourism entails MN. And the Weather Watchers example, as I understand it, is clearly intended to support an argument against this thesis.

[i]t may be noted that the point of the present story would be unaffected if one supposed that there were observable changes in the Weather Watchers that were not instances of intentional behavior but that could nevertheless be supposed to constitute evidence for their having mental lives (Strawson 1994, p257).

For Strawson, the relevant feature of the case is that the Weather Watchers cannot intentionally move their bodies or be disposed to do so. It is irrelevant whether their bodies can undergo unintentional movements, say if they have reflexes. So the Weather Watchers are supposed to lack exactly the kind of disposition that MN says they have if they have a desire. Hence they are a counter example to MN.

To this point there is nothing in the example beyond a straightforward denial of MN. But there is a further feature that Strawson highlights about the Weather Watchers that the example seems to turn on: the Weather Watchers do not have intentions (Strawson 1994, p252). It is this feature that he says makes the example 'potentially more revealing' (Strawson 1994, p252) about the connection between desire and action than a case where someone cannot actually act but can have an intention to do so. Although he does not state it clearly, the idea seems to be that having an intention to act is a necessary

condition for having a disposition to act.<sup>40</sup> So if the Weather Watchers can have desires yet cannot have intentions to act in ways that they believe will bring about what they desire then, in virtue of this they cannot have a disposition to act in ways they believe will bring about what they desire. Thereby they are a counter example to MN as MN claims that having a desire entails having such a disposition to act. Rather than being just the simple argument above, Strawson's objection is like the following extended argument (this is what I will call Strawson's objection hereafter):

- 1) If MN is true then it is not possible that *S* desires that *p* and *S* is not disposed to act in ways *S* believes will bring about *p*.
  - 2) If *S* is disposed to act in ways *S* believes will bring about *p* then *S* can have an intention to act in a way *S* believes will bring about *p*.
  - 3) It is possible that *S* desires that *p* and that *S* cannot have an intention to act in a way *S* believes will bring about *p*.
- C) So MN is false.

### 2.3 – Options.

---

<sup>40</sup> Strawson uses the expression 'trying to \_' but it is plausible from the text that he means by this 'intending to \_' (Strawson 1994, p275).

The argument is valid. But by making explicit the additional premises in Strawson's objection in this way we see that there are in fact two ways for someone to respond in defense of MN: they could either deny premise 2 or deny premise 3. Strawson gives the Weather Watchers example in support of premise 3. In the following sections I will argue that the example fails to do so, so will deny that premise. But what about premise 2, why should we accept this?

In fact this premise is very plausible: it seems to follow from the way that action is understood here, in terms of intentional bodily movement. Consider the following example from Davidson:

Suppose that a man boards an aeroplane marked 'London' with the intention of boarding an aeroplane headed for London, England [...]. [This] explains why he intentionally boarded the plane marked 'London'. As it happens, the plane marked 'London' was headed for London, Ontario, not London, England, and so [it] cannot explain why he boarded a plane headed for London, England. [It] can explain why he boarded a plane headed for London, Ontario, but only when [it is] conjoined to the fact that the plane marked 'London' was headed for London, Ontario; and of course [it] cannot explain why he intentionally

boarded a plane headed for London, Ontario, since he had no such intention (Davidson 1980, p84).<sup>41</sup>

As I understand him Davidson is arguing that to do something intentionally requires that one have an intention to do that particular thing. Note two important points here. First, having such an intention does not necessarily involve having a conscious, mental event that is the intention and that is accessible to introspection. In this respect intentions are most plausibly thought of like beliefs, desires and other kinds of propositional attitudes. There can be conscious intentions, just as there can be conscious beliefs. But it is not necessary for someone to have a belief that *p* that they have a particular mental event that is this belief. I might believe that Bugs and Roger are both rabbits without ever consciously entertaining that belief. And even if I do have that belief consciously on occasion, other things being equal I continue to believe it between different episodes of consciously entertaining it, at those times when I am not doing so. The same seems plausible in the case of intentions, *mutatis mutandis*.

What kind of mental state is an intention? There are currently two standard ways of characterizing intentions, either reductively or non-

---

<sup>41</sup> From essay 5: *Intending*, in Essays on Actions and Events (1980).

reductively. Davidson gives the first kind of account when he says the following:

An action is performed with a certain intention if it is caused in the right way by attitudes and beliefs that rationalize it.

If this account is correct, then acting with an intention does not require that there be any mysterious act of the will or special attitude or episode of willing (Davidson 1980, p87).

By 'attitudes and beliefs' here he means desires and belief: indeed, Bratman calls it the 'desire-belief model' when he describes reductive accounts like Davidson's (Bratman 1984, p 375). According to these accounts, having a particular intention is constituted by having a particular complex of desire(s) and belief(s).

On the other hand there are non-reductive accounts of intentions such as Bratman's own. According to non-reductive accounts intentions are a sui generis kind of mental state: having an intention is not constituted by having some complex of other kinds of states but rather by having a distinct kind of state that typically stands in certain relations. Bratman tries to illustrate this in the context of peoples' actions over time. These extended periods of acting can be thought of as following plans that

help guide our later conduct and coordinate our activities over time in ways in which our ordinary desires and beliefs do not. Intentions are typically elements in such coordinating plans. Once we recognize this central role intentions play in our lives the natural view to take, I think, is that intentions are distinctive states of mind, not to be reduced to clusters of desires and beliefs (Bratman 1984, p376).

But even though intentions are not reducible to desires and beliefs on this account, they have characteristic relations with desires, beliefs, other kind of mental state, and actions. In light of this it is plausible that they are a state of the same general kind as desires and beliefs: that is intentions are a kind of propositional attitude. It is just that they are a distinct kind of propositional attitude, in the way that a hope is a distinct kind of propositional attitude from a belief or a desire. So on either reductive or non-reductive views, it is plausible that having a particular intention is a matter of having a particular propositional attitude or set of propositional attitudes. We typically do not think that there must always be some mental event or conscious state when we have other kinds of propositional attitudes, so there is no reason to think this is necessary when having an intention.

A second point to note is that what is necessary for doing something intentionally is having an intention to do that particular thing. This is an intention *de re*, not an intention *de dicto*.<sup>42</sup> In addition, this intention must be causally connected in the right way with the movement.<sup>43</sup> So the intention that I must have for the event of my arm rising to be an action, an intentional bodily movement, is an intention that *I raise my arm*. It is not necessary that I have an intention that *I act to raise my arm*, or to paraphrase, an intention that *I raise my arm by having an intention that I raise my arm*.

In light of these considerations premise 2 seems very plausible. It says that it is necessary for *S* to have a disposition to act in a particular way, *m*, that *S* can have an intention to act in that way, an intention that *S* does *m*. And recall from the Introduction, having a disposition to act is a matter of acting just in case a particular set of conditions obtain. That is, saying that *S* is disposed to act is to say that for a particular set of propositions, *x, y, z*, if *x, y, z*, are true then *S* will act. For the purposes of this dialectic, acting in a particular way entails intentionally moving your body in a particular way. And as we

---

<sup>42</sup> The alternative, that it is necessary for doing something intentionally that someone has an intention to do it *de dicto*, is not plausible. It suggests that someone would have to have the concept of action, or the concept of having an intention as a necessary condition for acting. But conceptually unsophisticated creatures, like young children and animals, can act. So it is not necessary to have these concepts.

<sup>43</sup> Treating intentions as propositional attitudes raises a number of complicated issues. For example, it implies that the objects of intentions are propositions. Yet, it is controversial whether attributions of intentions that are not expressed in the propositional form can be paraphrased into the propositional form without changing the meaning of the attribution. An analogous problem comes from taking the objects of desires to be propositions, as discussed



have seen, intentionally moving your body in a particular way is a matter of having an intention to move in that way that causes the movement (in the right way). So to say that *S* is disposed to do action *m* is to say that for a particular set of propositions, *x, y, z*, if *x, y, z*, are true then *S* will do *m* caused (in the right way) by *S* having an intention that *S* does *m*. But now premise 2 seems trivial. It says that it is necessary that if there is a particular set of propositions *x, y, z*, such that if *x, y, z*, are true then *S* will do *m* caused (in the right way) by *S* having an intention that *S* does *m*, then *S* can have an intention that *S* does *m*.

However, there are some who might deny premise 2: those like Hursthouse, for example, might do so. But if so this would be because she rejects Davidson's account of intentions in particular. According to Davidson, having an intention to act in a particular way consists in having a complex of desire(s) and belief(s) that rationalizes the action. Hursthouse claims that there can be cases where someone acts, they do something intentionally, yet they do not have such a complex of desire(s) and belief(s) that rationalize the action.<sup>44</sup> She gives examples like 'rumpling the hair of, or generally messing up the person or animal one loves; talking to her photograph as one passes it,' and 'violently destroying or damaging anything remotely connected with the

---

in the Introduction. However, as I am doing for desire, I will not attempt to address such problems here: rather, I will assume that there is some solution available.

<sup>44</sup> Hursthouse, R. 1991: 'Arational Actions', *The Journal of Philosophy*, Vol.88, No.2 (Feb., 1991), pp. 57-68.

person (or animal, or institution)' one's [anger or hatred] is directed toward, e.g. her picture, letters or presents from her' (Hursthouse 1991, p58). These are cases where someone clearly does something intentionally but where they cannot plausibly be attributed some beliefs or desires that would rationalize what they do. If Davidson's account of intentions is correct then these are cases in which someone does something intentionally yet lacks an intention to do it. So they would be counter examples to premise 2.

As we have seen, however, Davidson's account is only one of a number of accounts of intentions. So even if Hursthouse is right, that it is possible to act, to intentionally move your body, without having some combination of desire(s) and belief(s) that rationalizes the action this does not show that it is possible to intentionally move your body without having an intention to do so. What these cases show is that either having an intention to do *m* is not necessary for intentionally doing *m* or that having an intention to do *m* is not constituted by having a complex of desires and beliefs that rationalizes doing *m*: that either premise 2 is false or Davidson's account of intentions is false. So Hursthouse's examples do not show that we should reject premise 2. In light of what both Strawson and I mean by acting here, and the plausibility of that premise in relation to this I will accept it for the purposes of this discussion.<sup>45</sup>

---

<sup>45</sup> I will discuss a different aspect of this issue, the relationship between having a disposition to act and having an intention, in the dissertation's Conclusion.

## 2.4 – Two Versions of the Weather Watchers.

Returning to the main argument, I will defend MN by denying premise 3 of Strawson's objection. Recall, premise 3 says the following:

- 3) It is possible that *S* desires that *p* and that *S* cannot have an intention to act in a way *S* believes will bring about *p*.

If we accept premise 2 of the objection then the third premise seems to amount to a straightforward denial of MN. So why should we accept it?

Strawson attempts to support premise 3 with the Weather Watchers example. But the example is under-described in an important way: he says that the Weather Watchers do not have intentions, but is this a necessary fact about them or merely contingent? We can distinguish two cases: first, there is Wayne the Weather Watcher, who does not have intentions necessarily. Wayne cannot have that kind of mental state, perhaps in a similar way that someone who is blind cannot have visual perceptions, or someone who is a complete amnesiac cannot have memories. In the next section I will argue that Wayne cannot have desires either, so is not an example of someone who has a

desire and lacks an intention to act. If this is right then Wayne does not support premise 3 and is not a counter example to MN.

Second, there is Dwayne the Weather Watcher, who never actually has an intention but where this is just a contingent fact about Dwayne. Unlike Wayne, it is not necessary that Dwayne does not have intentions. Clearly Dwayne does not support premise 3 of Strawson's argument as Dwayne can have an intention to act. Moreover, it is plausible that if Dwayne has a desire that  $p$  then Dwayne does have a disposition to act in ways that Dwayne believes will bring about  $p$ . Suppose that Dwayne desires that he has an apple and he believes, falsely, given that he is a Weather Watcher and cannot actually move, that he can get an apple by raising his arm to the apple tree above him. According to MN, Dwayne is disposed to raise his arm to the apple tree. Recall, from section 2.3, to say that  $S$  is disposed to do action  $m$  is to say that for a particular set of propositions,  $x, y, z$ , if  $x, y, z$ , are true then  $S$  will do  $m$  caused (in the right way) by  $S$  having an intention that I do  $m$ . So according to MN if Dwayne has this desire that he has an apple there is a particular set of propositions such that if they are true then Dwayne will raise his arm to the apple tree caused (in the right way) by Dwayne having an intention that he raises his arm to the apple tree. As I have discussed, it is a difficult task to specify which propositions will be in this set, and in this case it is likely that some of them would require large scale changes in Dwayne for

them to be true.<sup>46</sup> One of the propositions in this set will be the proposition that Dwayne has an intention that he raises his arm to the apple tree. But in Dwayne's case this particular proposition might easily be true without him undergoing any significant changes that might make us doubt that the person that acts is the same person, Dwayne the Weather Watcher, that the disposition to act is attributed to.<sup>47</sup> From the way that he describes the original example, highlighting the possibility of having an intention to act, Strawson seems to suggest that whether or not someone can have an intention to act is the crucial fact about them that determines whether they can have a disposition to act. He seems to consider other facts, such as whether or not their bodies can actually move as unimportant to whether they can have such a disposition. So by Strawson's own lights it is plausible that if Dwayne has this desire that he has an apple then he does have a disposition to act in a way that he believes will bring about that he has an apple. So Dwayne is not a counter example to MN.

---

<sup>46</sup> For instance, one of the ways that Dwayne would have to change is that his body would have to change physically so that it could actually act. However, Strawson does not seem to think that the fact that these kinds of changes would be necessary for a disposition to manifest make it problematic to attribute someone with that disposition (Strawson 1994, pp. 268-274).

<sup>47</sup> Another way of putting this is to say that in order to manifest the disposition and actually act Dwayne need not undergo any changes that might make us doubt whether he was still the same individual that we attributed the disposition to. Strawson claims that if it were necessary that Dwayne change to such an extent in order to actually act then he could not be attributed with the disposition to act (Strawson 1994, pp. 268-274).

## 2.5 – Wayne the Weather Watcher.

Wayne the Weather Watcher cannot have an intention to act: this is a necessary fact about him. If Wayne can have a desire then Wayne is an example that does support premise 3: his case would show that it is possible that someone can have a desire that  $p$  but cannot have an intention to act in a way that he believes will bring about  $p$ . As it is a necessary condition for Wayne to be disposed to act in this way that he can have this intention then he would be a counter example to MN. However, if Wayne cannot have an intention then it is plausible that he cannot have a desire. Recall from section 2.3, according to the standard accounts of intentions they are either reducible to a complex of desire(s) and belief(s) or they are not reducible. If a reductive account is true having an intention consists in having certain desires and beliefs. So if Wayne cannot have any intentions then this must be because he cannot have either desires or beliefs, or cannot have both together. Each of these alternatives is *prima facie* equally plausible, and according to at least the first one Wayne cannot have desires.<sup>48</sup> If that is the reason then the example

---

<sup>48</sup> Note that Strawson says explicitly that the Weather Watchers can have beliefs. So he would not claim that the fact that Wayne cannot have intentions here is because he cannot have beliefs. Moreover, it is plausible that if Wayne cannot have beliefs then he cannot have desires either (and conversely that if he cannot have desires then he cannot have beliefs). Current consensus is that both beliefs and desires should be defined functionally, in terms of their characteristic relations. And the relations with the other kind of state are among the paradigm relations of both beliefs and desires. So someone who cannot have one of these kinds of state could not have the other kind either.

does not support premise 3 of Strawson's objection: Wayne is not obviously an example of someone who can have a desire but cannot have an intention so does not show that this is possible. It is just as plausible that Wayne cannot have a desire as that he can. And if he cannot have a desire then it is irrelevant to MN that he cannot have an intention and cannot have a disposition to act.

But what about if a non-reductive account of intentions is correct? Even on this kind of account it is plausible that if someone cannot have an intention then they cannot have a desire. According to non-reductive accounts of intentions they are a *sui generis* kind of state, not reducible to a complex of other states like beliefs and desires. Nonetheless, as we saw from Bratman, according to these accounts intentions have characteristic relations with beliefs, desires, and actions, for example. But now recall from Chapter 1, the common consensus is that functionalism is the correct approach to analysing desire. This approach gives an account like the following (although most likely involving many more relations and properties):

*S* desires that *p* iff:

$[\exists(R1, R2, R3, F) \text{ If } S \text{ } R1\text{'s that } p \text{ then } S \text{ is disposed to } R2 \text{ in a way } S \text{ } R3\text{'s will bring about } p \text{ and } S \text{ has } F \text{ if } S \text{ } R3\text{'s that } p] \text{ and } [S \text{ } R1\text{'s that } p].$

What the relations and properties  $R1$ ,  $R2$ ,  $R3$  and  $F$  are will vary between versions of functionalist account of desire. But if a non-reductive account of intentions is correct then some of them will be relations of having an intention to do something. And as desire is defined in terms of these relations according to functionalism, if someone cannot have the relations then they cannot have a desire. So if someone cannot have an intention then they cannot have a desire. Hence, even on a non-reductive account of intentions Wayne cannot have desires, so is not a case that supports premise 3 of Strawson's objection.

If Wayne cannot have an intention then, whether a reductive or non-reductive account of intention is correct, it is plausible that he cannot have desires. So Wayne does not support premise 3: he does not show that it is possible that there can be someone that can have a desire yet that cannot have an intention. But neither does Dwayne support premise 3: he can have a desire but he can also have an intention. And Wayne and Dwayne seem to exhaust the possible ways of understanding the Weather Watchers example. So it fails to support Strawson's objection to MN.

## **2.6 – Shane the Weather Watcher.**



In fact there is another way of reading the Weather Watcher's example although this departs slightly from what Strawson actually says when describing it. Recall, Strawson says that the Weather Watchers have desires but do not have intentions (Strawson 1994, p252). But all that is needed to support premise 3 is an example of someone who has desires yet cannot have intentions to act. So, consider Shane the Weather Watcher who cannot have an intention to act. With respect to having this kind of intention Shane is like Wayne is with respect to having any kind of intention. Shane cannot have intentions like an intention that he raises his arm, or an intention that he opens the door. However, Shane can have other kinds of intentions. For example, Shane can have intentions like an intention that he will be as good as he can be, or an intention that he does not kill his plant.<sup>49</sup> With respect to having intentions that are not intentions to act Shane is like Dwayne is with respect to having any kind of intention.

Compare for illustration an analogous case with beliefs about a particular kind of thing: suppose I cannot have beliefs about fresh fruit but can have all other kinds of belief. For example, I can have beliefs like a belief that Canberra is East of Sydney, a belief that kicking cats is wrong, a belief that I am over six feet tall, and so on. But I cannot have beliefs like a belief that

---

<sup>49</sup> This assumes that having this kind of intention does not reduce to or entail having an intention to act. If that were the case then Shane would be exactly like Wayne: Shane could not have any intentions at all. In which case Wayne and Dwayne would exhaust the possible readings of the Weather Watchers example and there would be no counter example to MN.

apples grow on trees, or a belief that mangos and bananas are the same colour, or a belief that tomatoes have gone up in price recently. No propositions that make reference to fresh fruit can be the object of my beliefs. With respect to having intentions to act Shane is like me with respect to having beliefs about fresh fruit.

If Shane is possible then premise 3 of Strawson's objection is true: Shane is an example of someone who can have a desire that *p* but cannot have an intention to act in a way that he believes will bring about *p*. If he cannot have such an intention then, from premise 2, he cannot have a disposition to act in a way that he believes will bring about *p*. But MN entails that he will have this disposition if he has a desire that *p*, so Shane is a counter example to MN.

## **2.7 – Responses to Shane.**

How can we respond in defense of MN? Note that we cannot straightforwardly deny that Shane can have desires as we could with Wayne. As discussed already, according to functionalist accounts of desire having a desire consists in being in a state that has a certain set of relations, and among these are relations of having an intention. Wayne cannot have any intentions

so there are no states that Wayne can be in that have the set of relations necessary for being a desire. But Shane can have at least some relations of having an intention, those that are not relations of having an intention to act in particular ways. In practice, functionalists typically allow some flexibility here:<sup>50</sup> the set of relations that a particular account might give as an analysis of a particular kind of state will typically be large and complex. People usually concede that it is not necessary to have all of the relations in that set in order to have that state. Provided that someone has enough of them then we will typically say that they have that kind of state. Of course, there are difficult questions about how many is enough, and which ones are most important to have. If, in fact, intentions to act make up a significant proportion of these relations, and they are central to the analysis of desire then it will be necessary for having a desire that someone can have intentions to act. In which case Shane, like Wayne, cannot have desires and fails to be a counter example to MN for the same reason as Wayne. But if intentions to act are more peripheral to the analysis of desire then the fact that Shane cannot have them will not prevent him having desires. This turns on the details of the correct functionalist account of desire, something that it is difficult to predict in

---

<sup>50</sup> See for example, (Braddon-Mitchell and Jackson 1996, pp. 46-48) for discussion. They say, '[t]here is a list of features that we regard as paradigmatic of [the mental state being analysed]...But nothing in that list is sacrosanct...What matters is that enough of the list is satisfied or near enough satisfied, and what counts as enough may itself be a vague matter, and may change with time. But enough of the list must be satisfied or near enough satisfied' (pp. 47-48).

advance. And whereas it is very plausible that Wayne cannot have desires however this turns out, it is less obvious that Shane will be the same.

Instead we should respond by straightforwardly denying that Shane is possible. Notice how closely this case has to be specified so that Shane has exactly the properties described in premise 3 yet so that the problems for the examples of Wayne and Dwayne can be avoided. But premise 3 is effectively a denial of MN. So in an argument against MN it is not a legitimate argumentative step to support premise 3 with an example that has been specified so precisely to show that the premise is true unless there is independent motivation to think that the example is possible. Otherwise saying that the example is possible is merely asserting that MN is false. Compare the analogy with belief from above: the case where I cannot have beliefs about fresh fruit but can have other kinds of belief. If we tried to argue against a particular view about belief with this case then there would have to be some reason to think that it is actually possible that is independent of thinking that that theory was false. The case is an unusual and unfamiliar one, so someone who held that theory of belief would be justified in demanding such a reason. Otherwise the example would beg the question against that theory. The situation is similar here with Shane and premise 3 of Strawson's objection. Without independent reason to think that Shane is possible then we should simply deny that he is. But in the following sections I will argue that

the only obvious motivations are not convincing.<sup>51</sup> So in the absence of a further argument to support it the example is implausible.

## 2.8 – Motivation for Shane.

What reasons are there for thinking that someone like Shane is possible other than already thinking that MN is false? What other features of the example are there that could make it plausible that Shane cannot have intentions to act despite being able to have intentions of other kinds? There do not appear to be any psychological facts about Shane that can do this. According to Strawson the Weather Watchers are like us psychologically: they can have the same kinds of mental states as we can, aside from intentions, and they are conceptually sophisticated. For example, Weather Watchers can have beliefs about acting in particular ways and desires about acting in particular ways, so Shane can have the appropriate concepts for having intentions to act. However, the Weather Watchers are unlike us physically: they cannot actually act in any particular ways because they cannot intentionally move their bodies.

---

<sup>51</sup> There are difficult and controversial issues here concerning the legitimacy of denying the possibility of a putative counter example as a response to an argument. However, I will not attempt to engage with those here. See e.g. D. Stoljar, 'Two Conceivability Arguments Compared' (forthcoming in *Proceedings of the Aristotelian Society*) for further discussion of such issues in the context of comparing the perfect actor argument against behaviourism and the zombie argument against physicalism.

Does this physical fact about Shane, that he cannot act, make it plausible that he cannot have intentions to act but can have other kinds of intentions?

Note that there are analogies of this kind of relation: according to certain plausible and widely accepted theories there are particular kinds of mental state that someone can only have if certain contingent facts are true. For example, if externalism about content, and certain causal theories of meaning are true then someone can only have thoughts about water if they have come into the right kind of causal contact with samples of H<sub>2</sub>O.<sup>52</sup> It is because Twin Oscar has not had this kind of relationship with samples of H<sub>2</sub>O that his 'water' thoughts are not about water. Rather they are about twin water, that is XYZ, which is the substance in the lakes and rivers on Twin Earth and that he has this kind of causal contact with. Similarly, Jackson's example of Mary in the black-and-white room suggests that someone can only have thoughts that employ phenomenal concepts, like visualizing what red looks like, if they have had the right kind of phenomenal experiences.<sup>53</sup> Before her release Mary cannot say, visualize what red looks like, and perhaps

---

<sup>52</sup> See e.g. Putnam, H. 1975: 'The Meaning of Meaning', *Language, Mind and Knowledge. Minnesota Studies in the Philosophy of Science*, vol. 7, ed. Keith Gunderson (Minneapolis: University of Minnesota Press), pp. 131-193. Repr. in *Mind, Language and Reality* (1975), pp. 215-271.

<sup>53</sup> See e.g. Jackson, F. 1982: 'Epiphenomenal Qualia', *Philosophical Quarterly*, 32, 127, pp. 127-136, (April, 1982), and 1986: 'What Mary didn't Know', *Journal of Philosophy*, 83, 5, pp. 291-295, (May, 1986).

as a consequence cannot know what it is like to see red, because she has not had any red experiences.

But note also that in these cases it is some event that is missing from the history of Twin Oscar and Mary that prevents them from having the respective mental state. At time  $t$  Twin Oscar cannot have thoughts about water because he has not come into the right kind of causal contact with samples of H<sub>2</sub>O before  $t$ . The obvious suggestion for an analogous event that is missing from Shane's history that is necessary for having an intention to act at time  $t$  is that he has actually acted before  $t$ . Is it plausible that someone can only intend to act in a particular way at  $t$  if they have actually acted before  $t$ ?

What is not plausible is that it is necessary for having an intention to act in a particular way that someone has previously acted in that specific way. It is not plausible, for example, that at  $t$  someone can have an intention that I raise my arm only if at some time before  $t$  they have actually raised their arm. This would entail that someone could not intend to perform a particular movement unless they had previously performed that exact movement. But clearly it is possible to think of some movement that you have never actually performed before and intend to do it. Suppose that before now I have never wiggled my left little finger yet have wiggled each of my other fingers at some time or other. Clearly I could now have an intention that I wiggle my left little

finger. So it is not necessary for having an intention to act in a particular way that someone has previously acted in that exact way.

Rather, the suggestion is that someone can have an intention to perform a particular action only if they have previously acted in some way or other. Call this claim (IA). For a particular action  $m$ , (IA) says the following:

(IA)                    If  $S$  at  $t$  can have an intention to  $m$  then there is an action  $n$  such that  $S$  at  $t-1$  has  $n$ -ed.

According to (IA) if I can now have an intention that I raise my arm then at some time before now I have acted in some way or other: before now I have wiggled my fingers or lifted my leg, or walked to the door, or intentionally moved my body in some other way. If (IA) is true then there is independent motivation for thinking that a case like Shane the Weather Watcher is possible, and Shane can be used to support premise 3 of Strawson's argument.

## **2.9 – A Regress of Intentions and Actions.**

Unfortunately for Shane the Weather Watcher, (IA) is false: it is not necessary for having an intention to act in a particular way that you have previously



acted in some way or other. If it was true then it would not be possible to perform a first action. There would be a regress of what is necessary for having an intention to act and what is necessary for acting. The regress develops in the following way:

- 1) If  $S$  at  $t$  can have an intention to  $m$  then there is an action  $n$  such that  $S$  at  $t-1$  has  $n$ -ed [from (IA)]
- 2) If there is an action  $n$  such that  $S$  at  $t-1$  has  $n$ -ed then  $S$  at  $t-1$  can have an intention to  $n$ .
- 3) If  $S$  at  $t-1$  can have an intention to  $n$  then there is an action  $o$  such that  $S$  at  $t-2$  has  $o$ -ed [from (IA)]
- 4) If there is an action  $o$  such that  $S$  at  $t-2$  has  $o$ -ed then  $S$  at  $t-2$  can have an intention to  $o$ .
- 5) If  $S$  at  $t-2$  can have an intention to  $o$  then there is an action  $q$  such that  $S$  at  $t-3$  has  $q$ -ed [from (IA)]

And so on ad infinitum.

The problem arises because of what both Strawson and I mean here by acting, that is intentional moving your body. As discussed in section 2.3 an intentional bodily movement is a bodily movement that is caused in the right way by having an intention to make that movement. Steps 2 and 4 in the

regress are just examples of this. But (IA) says that it is necessary for having an intention that someone has previously acted, that is that they have previously made an intentional bodily movement. This is steps 1, 3, and 5. Together these entail that someone can have an intention to act only if they have already acted and that they can have already acted only if they can have an intention to act: no one could perform a first action, and consequently no one could act. Obviously people do act, so either (IA), and steps 1, 3 and 5, is false, or acting does not require intending to act, and steps 2 and 4 are false. But this second claim, and steps 2 and 4, seems to follow naturally from what we mean by acting: indeed it seems to follow from premise 2 of Strawson's objection. So there would be little point for Strawson in rejecting it to save (IA) so as to motivate an example that is supposed to support premise 3 of his objection. Moreover, as I discussed in section 2.3 it is a familiar claim in philosophy, and it is *prima facie* more plausible that (IA). So it is (IA) that we should deny here.

If (IA) is false is there any other independent reason to accept that someone like Shane is possible? As I understand it there are no other facts about the Weather Watchers, as Strawson describes them that make it plausible that there could be a Weather Watcher with the psychological limitations that Shane has. So there is no reason to accept that a case like Shane is possible unless you already think that premise 3 of Strawson's

objection is true: the example cannot support that premise. But neither Wayne nor Dwayne, the other versions of the Weather Watchers example support that premise either. So Strawson has not given a successful counter example to support his objection to MN.

## **2.10 – Conclusion.**

According to MN if someone has a desire then they are disposed to act in ways they believe will bring about what they desire. Galen Strawson attempted to show that MN is false with his example of the Weather Watchers. Strawson appealed in his objection to the relation between having a desire and having an intention, and the relation between having an intention and having a disposition to act. He claimed that necessarily someone can have a disposition to act only if they can have an intention to act. And he gave the Weather Watchers example to show that it is possible that someone can have a desire but cannot have intentions. This would entail that MN was false. However, Strawson's example fails to show that this is possible. The case is under-described in crucial respects. Moreover I have argued that none of the ways in which the example can be adequately described without begging the question against MN are a counter example to it. Either it is a necessary fact about the

Weather Watchers that they cannot have intentions, in which case they cannot have desires, or it is only a contingent fact about them that they do not have intentions, in which case if they have desires then they do have dispositions to act. So Strawson's objection fails.

Nonetheless, the objection raises an interesting issue: how are desires, intentions and dispositions to act related to one another? Strawson claimed that there is a necessary connection between dispositions to act and intentions. And he denied that there is a necessary connection between desires and intentions. While it does seem that there is some kind of connection between dispositions to act and intentions has Strawson characterized it correctly? And if, as I have argued, there is a necessary connection between desires and intentions might this provide an explanation of the connection between desires and dispositions to act? These are interesting questions and I will return to them briefly in the final Conclusion, but I will not attempt to address them in detail in this dissertation.

# CHAPTER 3

## TROUBLESOME DESIRES.

### 3.1 – Introduction.

MN is a thesis about all things and all desires. It says that if any thing has any kind of desire then it is disposed to act in ways it believes will satisfy that desire. In the previous chapter I discussed an objection to MN made by Galen Strawson. Strawson attempted to show that MN is false by giving an example of a certain kind of thing, the Weather Watchers that could have desires and not be disposed to act. I argued that his example failed to show this, that either the Weather Watchers cannot have desires, or that if they do have desires then they do have dispositions to act.

In this chapter I will discuss objections that are directed at the other universally quantified aspect of MN: these objections claim there are particular kinds of desire that someone can have without being disposed to act. In these cases it is the specific content of the desire, what it is a desire for, that makes them counter examples MN.

These troublesome cases for MN are typically of one of three kinds: first, desires for something the desirer wants without having to get it through their own actions. Call these ‘self-passive desires’: they are desires like a desire that your child passes an exam on their own. Second, desires for things the desirer believes are necessarily the case. Call these ‘necessity desires’: they are desires like a desire that there are infinitely many prime numbers. Third, desires for thing the desirer believes are necessarily not the case. Call these ‘impossibility desires’: they are desires like a desire that I have a square circle. If someone can have desires like these and not be disposed to act in ways they believe will satisfy them then they are a counter example to MN: MN says that, for all  $p$ , if someone has a desire that  $p$  then they will be disposed to act in ways they believe will bring about  $p$ .

However, I will argue that none of these kinds of desire succeed as counter examples. For most of the chapter I will discuss self-passive desires. I argue that these desires only appear to be counter examples to MN because people have misunderstood their content. Once we see clearly what the object of these desires really is they are no longer a problem for MN.

At the end of the chapter I will more briefly discuss necessity desires and impossibility desires together. I argue again that these desires appear to be counter examples to MN because of a misunderstanding of their content. But in these cases this misunderstanding is an artifact of this kind of content

specifically: it is because their propositional objects are propositions that are necessarily true or necessarily false. There is a general problem concerning how it is possible to represent propositions like this: it is not specifically a problem for the case of desire. But we can represent these propositions so there must be a solution to the general problem. It is plausible that once we have this solution it will also apply to the specific case of desire: it will allow us to understand what the objects of necessity desires and impossibility desires really are. Once we see this clearly it is plausible that these desires will no longer appear to be counter examples to MN.

### 3.2 – Self-Passive Desires.<sup>54</sup>

According to MN, if *S* desires that *p* then *S* is disposed to act in ways that *S* believes will bring about *p*. This is a thesis about all kinds of desire, that is, it is supposed to be true of someone if they have any desire no matter what its object is. However, there is a common and seemingly reasonable kind of desire, for a particular kind of thing, where it appears that people who have these desires are not disposed to act as MN says. Timothy Schroeder describes the objects of these desires as

---

<sup>54</sup> Thanks to Daniel Friedrich for extensive discussions about this kind of counter example.

ends that the agent hopes to obtain but that the agent would not be willing, under any circumstances, to cause to obtain. Suppose I desire that a committee makeup its mind in my favour without my intervention. This is a state of affairs I might want very much, yet because of the very nature of the desire it makes no sense to try to act so as to satisfy it. What I want is that the committee make a certain decision *without* my needing to do anything (Schroeder 2004, p17).

Call desires of this kind 'self-passive desires'. Schroeder's example is a desire that the committee chooses me without my needing to do anything. That is, a desire that the committee chooses me and that I do not act to bring it about that the committee chooses me. Why is it a counter example to MN? It seems that what Schroeder wants here is that the committee chooses him. According to MN if he desires that the committee chooses him then he is disposed to act in ways he believes will bring about that the committee chooses him. But it seems that when Schroeder has this desire he is not disposed to act in any way that he believes will bring about that the committee chooses him. According to Schroeder he cannot sensibly be disposed to act if he has this kind of desire: if he manifested this disposition and acted in a way that he believed would bring about that the committee chooses him then in virtue of doing so he



would not get what he desires. He would not bring about that the committee chose him and that he did not act in a way he believed would bring this about. So this seems to be a desire for which  $S$  desires that  $p$  and  $S$  is not disposed to act in ways  $S$  believes will bring about  $p$ . If so then it is a counter example and MN is false.

### 3.3 – Options.

How can we respond in defense of MN? One response is to deny that when Schroeder has this desire he is not disposed to act. According to this flatfooted response all the case shows is that Schroeder has a desire and does not act. But this does not show that he is not disposed to act. Note that this is similar to the kind of responses I gave in the previous chapter to the different versions of the Weather Watchers example. Is it convincing here?

It is difficult to adjudicate in this kind of disagreement where there is a conflict of intuitions about a particular example. However, there are a number of differences between the Weather Watchers cases and the examples of self-passive desires. When I denied that Wayne the Weather Watcher, who cannot have intentions, could have a desire I attempted to support this by arguing that it is implied by standard and plausible accounts of intentions. So I gave an

independent motivation for denying that Wayne was a counter example to MN. Conversely when I denied that Shane the Weather Watcher was possible I attempted to support this by arguing that it needed to be independently motivated because that kind of case is unfamiliar and it appeared to be ad hoc: it seems gerrymandered to have all and only those precise features necessary to be a counter example to MN. But I argued that there was no obvious independent motivation for it. Yet Schroeder's example of a self-passive desire is unlike the Weather Watchers cases in these respects. First, self-passive desires are not unfamiliar and do not appear to be ad hoc stipulations. Schroeder's own example, a desire that the committee chooses him without his needing to do anything to bring this about seems to be a familiar and reasonable desire that someone might have in the situation he describes. Also, there lots of everyday situations in which it would be reasonable for someone to have a self-passive desire. Consider some further examples: (a) Mrs. Cunningham desires that Joanie passes her high-school exams on her own. (b) Richie desires that Lori-Beth goes on a date with him because she is attracted to him for who he is. (c) Fonzie desires that Chachi chooses to work with him at the bike-shop of his own volition. (d) Mr. Cunningham desires that the Bucks win the NBA Championship play-offs in a fair contest. These are all familiar and reasonable desires for someone to have and are all examples of self-passive desires.

In addition, Schroeder does give independent motivation for saying that in this case when he desires that the committee chooses him without his needing to do anything to bring that about he is not disposed to act in ways he believes will bring about that the committee chooses him. Recall, Schroeder says that ‘because of the very nature of the desire it makes no sense to try to act so as to satisfy it’ (Schroeder 2004, p17). The idea seems to be that because this desire is a desire that he has something that he does not act to get himself then, if he were to act in a way he believed will bring about what he desires then, in virtue of doing so he would frustrate his desire. Moreover, if he understood what it is that he desired and had thought things through then he could not fail to recognize that by acting in that way he would frustrate that desire: he would also believe that by acting in that way he would not bring about what he desires. So if MN entails that someone who has this desire does have a disposition to act in a way they believe will bring about what they desire then MN entails that someone who has this desire is disposed to act in a way that they also believe will frustrate that very desire.<sup>55</sup> But according to

---

<sup>55</sup> In addition, MN seems to entail that if someone has a self-passive desire then they will have contradictory beliefs. If there is some action, *m*, that they believe they can perform that will bring about what they desire yet they will recognize that by doing *m* they will fail to bring about what they desire, if they think things through, then they will both believe that by doing *m* I can bring about *p*, and believe that by doing *m* I cannot bring about *p*. However it is the kind of irrationality discussed in the main text, that MN entails that in virtue of having a self-passive desire someone is disposed to act in ways that they believe will frustrate that desire that Schroeder seems to focus on in his discussion. And it is that kind of irrationality that appears more relevant to being rational in practical matters, concerning the relations between one’s beliefs, desires and actions. There are interesting questions about the relation between such practical rationality and what might be called theoretical rationality, concerning the relations between one’s beliefs alone, but I will not engage with them here.

Schroeder insofar as someone is rational they will not be like this: they will not knowingly frustrate their desire in virtue of having it. So, insofar as someone is rational they will not have the disposition that MN entails. It seems that if MN is true then no one who is rational can have a self-passive desire. But that is clearly false, as we can see from the examples: in each case there is nothing unreasonable in that person having a self-passive desire. So Schroeder does not merely assert that that his example is one in which someone has a desire and is not be disposed to act in a way they believe will bring about what they desire. He supports the claim by suggesting that what it is for someone to be rational implies that they will not have this disposition.<sup>56</sup> Of course, it then appears that Schroeder's explicit target with the example is not MN but rather the related thesis I distinguished in the Introduction, MN(rat). Recall, MN(rat) says that if *S* desires that *p* then *S* is disposed to act in ways *S* believes will bring about *p* insofar as *S* is rational. But MN(rat) is entailed by MN so this is a counter example to MN nonetheless.

---

<sup>56</sup> Is this in fact what Schroeder is claiming? That is, is the case supposed to be one in which someone has a desire and they do not have a disposition to act in a way they believe will bring about what they desire, and they are rational? Or is he claiming more than this: is the case supposed to be one in which someone has a desire and they do not have a disposition to act in a way they believe will bring about what they desire, and this is because they are rational? That is, is Schroeder claiming that it is being rationality that is somehow responsible for his lacking a disposition to act when he has this desire? If this is what Schroeder intend the example to show then it is not clear that it would be a counter example to MN(rat) or MN as neither thesis is committed to anything about what is responsible for the relation between having a desire and being disposed to act. This is another difficult issue to judge, exactly how Schroeder intends the example to be taken but for the purposes of this discussion I will assume he is making the weaker claim and that he does intend the case to be a counter example to MN(rat) and MN.

So it seems that the flatfooted response to Schroeder's case is not well supported here: it is not very plausible to respond by straightforwardly denying that in the example when Schroeder has the self-passive he is not disposed to act. However, we can sidestep the problems about arbitrating between conflicting intuitions: we can grant to Schroeder that in the example he is not disposed to act in ways he believes will bring about that the committee chooses him. Nonetheless, as I will argue in the next section the case fails to be a counter example to MN.

### **3.4 – Self-Passive Desires are not a Counter Example to MN.**

What would a counter example to MN be like? According to MN if  $S$  desires that  $p$  then  $S$  is disposed to act in ways  $S$  believes will bring about  $p$ . So a counter example would be a case in which someone desires that  $p$  and is not disposed to act in a way they believe will bring about  $p$ . But Schroeder's example is not like this. Recall, in his example Schroeder desires that the committee chooses him and that he does not act to bring about that the committee chooses him, and he is not disposed to act in ways he believes will bring about that the committee chooses him. But MN does not entail that he will have this disposition. This is because Schroeder's desire here is not a

desire that the committee chooses him. Rather, his desire is a desire that the committee chooses him and that he does not act to bring about that the committee chooses him. The content of his desire is a conjunction, that a particular state of affairs comes about and that it does not come about in a certain set of ways. MN entails that he is disposed to act in ways he believes will bring about this conjunction: he is disposed to act in ways he believes will bring about that the committee chooses him and he does not act to bring about that the committee chooses him. Schroeder says nothing about whether in this case he has this disposition, and it is irrelevant to MN that he lacks a disposition to act in ways he believes will bring about that the committee chooses him. So Schroeder's case is not a counter example to MN.

We can make this response more precise: consider again his example and the examples of other self-passive desires I gave in the previous section. They all have in common that the propositional object of the desire is a conjunction, that a particular state of affairs come about and that they do not act in a way that they believe will bring about this state of affairs. We can represent them all as a desire that *p and that I do not act to bring about p*. Call this conjunction '*q*'. According to MN if *S* desires that *q* then *S* is disposed to act in a way *S* believes will bring about *q*. In Schroeder's example he has a self-passive desire that *q* and he is not disposed to act in a way he believes will bring about *p*. And this is not a counter example to MN.

Moreover, in Schroeder's example, and the other examples of self-passive desires it is plausible that each person has the disposition that MN attributes to them. They are disposed to act in ways they believe will bring about  $q$ , or more precisely, they are disposed to act in ways they believe will bring about  $p$  and *I do not act to bring about  $p$* . Consider Mrs. Cunningham's self-passive desire that Joanie passes her high-school exams on her own. Making the object of her desire more explicit, Mrs. Cunningham desires that Joanie passes her high-school exams and that she does not bring about that Joanie passes her high-school exams. It is plausible that she is not disposed to act in ways she believes will bring about that Joanie passes the exams: she is not disposed to do things that she might believe will straightforwardly make it the case that Joanie passes her exam, such as provide Joanie with the answers to the exams beforehand, or bribe Joanie's teachers to mark her exam scripts leniently. But suppose Mrs. Cunningham believes that if she encourages Joanie to study then, although this will not make it the case that Joanie passes, it will make Joanie more likely to study and pass on her own. It is plausible that if Mrs. Cunningham has this belief and a self-passive desire that Joanie passes her high-school exams on her own then she is disposed to encourage Joanie to study. Hence, she has a desire that  $p$  and *I do not act to bring about  $p$*  and she is disposed to act in a way she believes will bring about that  $p$  and *I do not act to bring about  $p$* . So MN is true in this case.

Similarly, consider Schroeder's own example: Schroeder desires that the committee chooses me and he does not act to bring about that the committee chooses him. As he says, he is not disposed to act in ways he believes will bring about that the committee chooses him: he is not disposed to do things like bribe committee members to promote him to their colleagues, nor to spread rumours about the other candidates to undermine their applications. But suppose that he believes that if he wears a nice suit, prepares in advance for the interview, and gives firm handshakes to the committee members then, although this will not make it the case that the committee will choose him it will make them more likely to choose him of their own accord. If Schroeder does have this belief and a self-passive desire that the committee chooses me and he does not act to bring about that the committee chooses him then it is plausible that he is disposed to wear a nice suit, prepare in advance for the interview and give firm handshakes to the committee members. Like Mrs. Cunningham, in Schroeder's example he has a self-passive desire that *p* and *I do not act to bring about p* and he is disposed to act in ways he believes will bring about that *p* and *I do not act to bring about p*. This is just what MN entails. So these cases of self-passive desires, like Schroeder's are not counter examples to MN.



### 3.5 – Self-Passive Desires and Straightforward Desires.

Might there be a counter example to MN in a more complex case, closely related to Schroeder's example? Recall, a counter example would be a case in which someone has a desire that  $p$  and is not disposed to act in ways they believe will bring about  $p$ . In Schroeder's example he has a desire that  $p$  and *I do not bring about that  $p$*  and is not disposed to act in ways he believes will bring about  $p$ . So it is not a counter example to MN. But consider a different case: suppose the next interviewee on the shortlist, call him 'Schroeder\*' has a similar desire to Schroeder's, a self-passive desire that the committee chooses me and that he does not act to bring about that the committee chooses him. But suppose that in addition to this self-passive desire Schroeder\* also has what I will call the corresponding 'straightforward desire', a desire for the first conjunct of the propositional object of the self-passive desire alone. In this case Schroeder\*'s straightforward desire is a desire that the committee chooses him. So this is a case in which Schroeder\* has two desires: he has a desire that  $p$  and also a distinct desire that  $p$  and *I do not act to bring about  $p$* . According to MN Schroeder\* is disposed to act in ways he believes will bring about  $p$  and he is disposed to act in ways he believes will bring about  $p$  and *I do not act to bring about  $p$* . So if, as in Schroeder's example, Schroeder\* is not disposed to act in ways he believes will bring about  $p$  then this is a counter

example to MN: it is a case in which Schroeder\* has a desire that  $p$ , his straightforward desire that the committee chooses him, and he is not disposed to act in ways that he believes will bring about  $p$ , ways he believes will bring about that the committee chooses him. If this combination of desires and dispositions that Schroeder\* seems to have is possible then MN is false. The argument can be summarised as follows [where  $SD(p)$  stands for  $S$  desires that  $p$ , and  $SM(p)$  stands for  $S$  is disposed to act in ways  $S$  believes will bring about  $p$ ]:

- 1) If MN is true then it is not possible that  $SD(p)$  & not  $SM(p)$ .
- 2) It is possible that [ $SD(p)$  &  $SD(p \text{ and } I \text{ do not bring about that } p)$ ]  
& not  $SM(p)$ .
- C) MN is false.

### 3.6 – Options (part 2).

This argument is valid. So, to defend MN we have to deny premise 2, the premise supported with the example of Schroeder\*. Is it plausible that Schroeder\* has the combination of desires and dispositions required to show that premise 2 is true? He might fail to do so in at least two ways: he might

not have both the straightforward desire that the committee chooses him and the self-passive desire that the committee chooses him and that he does not act to bring about that the committee chooses him. Alternatively, if he has both desires he might actually be disposed to act in ways he believes will bring about that the committee chooses him. I will discuss the second option in the following section but first, in the rest of this section I will discuss the first option.

Could Schroeder\* have both the straightforward desire and the self-passive desire at the same time? One reason to think that he could do is if desire distributes over conjunction in the same way as belief is commonly taken to do. It is standardly assumed that belief distributes over conjunction, that is that if someone believes that *p and q* then this entails that they believe that *p* and they believe that *q*.<sup>57</sup> If desire is like belief in this respect then in virtue of having the self-passive desire this entails that Schroeder\* has both the straightforward desire and the self-passive desire together. A self-passive desire is a desire that *p and I do not act to bring about p*, so its propositional object is a conjunction. If desire distributes over conjunction then desiring *p and I do not act to bring about p* entails desiring that *p* and desiring that *I do not act to bring about p*. Indeed, if desire distributes over conjunction then

---

<sup>57</sup> See, for example Sorensen (1988), who says '[a]nyone who believes there is a non-trivial logic of belief is committed to the view that belief is orderly. The order is standardly described in terms of belief expanding rules and belief restricting rules. For example, the

there is no difference between Schroeder's original example and the case of Schroeder\*: Schroeder would have both a self-passive desire and the corresponding straightforward desire merely in virtue of having a self-passive desire.

However, in fact it is not plausible that desire distributes over conjunction in the same way as belief does. There are obvious cases in which someone desires to have two things together but does not want at least one of the things on its own. The propositional object of their desire is a conjunction but they do not desire each of the conjuncts individually. Suppose that Mr. Cunningham is a passionate supporter of the Milwaukee Bucks and desires that Lew Alcindor stays with the team and that he scores 30 points in the first game of the NBA Championship play-offs. The propositional object of his desire is a conjunction. Yet it is plausible that he does not desire each of the conjuncts individually. Suppose that his love of basketball is fervently one-eyed: he only likes the Bucks and despises all other teams and their players. It is plausible that Mr. Cunningham does not have a straightforward desire that Lew Alcindor scores 30 points in the first game of the NBA Championship play-offs. If Alcindor were to transfer to another team then he would not desire this: indeed he would more likely be strongly averse to it! So here Mr. Cunningham has a desire with a propositional object that is a conjunction but

---

principle that belief distributes over conjunction,  $B(p \ \& \ q) \supset (Bp \ \& \ Bq)$ , enables us to expand our list of beliefs given that the antecedent is on the list' (Sorensen 1988, p22).

does not have a desire for at least one of the conjuncts. So desire does not distribute across conjunction, and having a self-passive desire that *p and I do not act to bring about p* does not entail having both a self-passive desire that *p and I do not act to bring about that p* and a straightforward desire that *p*.

Nonetheless, it seems plausible as an empirical fact that someone could have both a self-passive desire and the corresponding straightforward desire together. A self-passive desire is a desire that a certain state of affairs come about in a certain way. But if someone has a desire like this it is plausible that they might also desire that the state of affairs itself come about, independently of how it comes about. Suppose in the earlier example that it has been an ambition of Schroeder\* for all of his adult life to get a job in that particular faculty. Perhaps it is in an extremely distinguished University, where Schroeder\* senior was an established Professor during his own career: Schroeder\* was raised according to the University's principles and extolling its virtues, and he is an alumnus of the University himself. It would make his life fulfilled if he got the appointment. But suppose also that Schroeder\* is less scrupulous than Schroeder. Of course, he wants to get the job in the right way so he has a self-passive desire that the committee chooses him and that he does not act to bring about that the committee chooses him. But he wants the job so much that he also straightforwardly desires that the committee chooses him, irrespective of how that decision is reached. So he has both a self-passive

desire and a straightforward desire together. This kind of situation seems obviously possible and it is plausible that there are actual cases like this. So it is not plausible to defend MN by arguing that the example of Schroeder\* does not show that the combination of desires described in premise 2 is possible.

### **3.7 – Two Dispositions.**

What about the second way that the example could fail to support premise 2: does the example show that it is possible for someone to have both a self-passive desire and a straightforward desire together and not be disposed to act in ways that they believe will satisfy the straightforward desire? It does not show this. In fact, it is plausible that in such a case they will have two dispositions to act, corresponding to each of their two desires, the self-passive desire and the straightforward desire.

If someone has a self-passive desire like a desire that the committee chooses me and I do not act to bring about that the committee chooses me then they have recognized that there are a number of ways in which the committee could make a particular decision. And what they want is for the committee to make that decision in only some of those ways. This is what it is to have this particular self-passive desire. As discussed above, it is plausible that they are

disposed to act in ways they believe will bring this about, for example they are disposed to wear a nice suit, prepare in advance for the interview and give firm handshakes to the committee members. But if, in spite of recognizing this they nonetheless have the corresponding straightforward desire, they want the committee to make that particular decision in any of the ways they could make it, then it is also plausible that they are disposed to act in ways they believe will bring about that the committee chooses them. Consider Schroeder\* once again. To make it seem plausible that it is a case in which he has both the self-passive desire and the straightforward desire, I had to embellish Schroeder's original example with respect to the importance to Schroeder \* of getting the position and his being somewhat unscrupulous. In light of this it seems plausible that he will be disposed to act in ways he believes will bring about that the committee chooses him. This does not mean that he will readily act in these ways. Perhaps his respect for the traditions and principles of the University also disposes Schroeder\* not to act in such ways. And this disposition may over-ride the disposition to act that he has in virtue of having the straightforward desire that the committee chooses me so that he does not act in ways he believes will bring about that the committee chooses him. But he has this disposition nonetheless. In which case there is no counter example to MN. Schroeder\* may have both a desire that *p* and *I do not act to bring about that p* and a desire that *p* but he also has both a disposition to act in ways

that he believes will bring about *p* and *I do not bring about that p* and a disposition to act in ways that he believes will bring about *p*.

This is not merely an artifact of this particular example. Rather, it seems that in cases where it is plausible that someone has both a self-passive desire and a straightforward desire then it will be just as plausible that they are disposed to act in ways that they believe will bring about what they straightforwardly desire as that they are not so disposed. Consider another case, ‘the unscrupulous heir’: suppose that Robert has always had a difficult relationship with his father. When he was younger he and his father frequently argued and he has never really felt that his father has any pride in, or affection for him. Suppose also that his father’s prize possession is a football shirt worn by Sir Bobby Charlton in one of the matches in the 1966 World Cup, signed by Charlton himself and given to Robert’s father in appreciation of the work he did in support of Charlton’s charity. Suppose that Robert’s father is on his death bed. The shirt is the only thing of value that his father owns and Robert thinks it would finally show that his father’s loves him if he left it to Robert in his will without him having to solicit it. So Robert has a self-passive desire that his father leaves him the shirt and that he does not act to bring about that his father leaves him the shirt. It is plausible that he is disposed to act in ways that he believes will bring about that his father leaves him the shirt and that he does not act to bring about that his father leaves him the shirt: for example, he



is disposed to visit his father in hospital, to bring him flowers and grapes, to reassure him that he is taking care of his home, and so on. Does Robert also have a straightforward desire that his father leaves him the shirt? It is easy to imagine a case in which this is plausible: suppose that he also has gambling debts to some vicious loan sharks and has no way of paying them off except by selling the shirt. So he has both a self-passive desire that his father leaves him the shirt and that he does not act to bring about that his father leaves him the shirt and the corresponding straightforward desire that his father leaves him the shirt. According to MN Robert is disposed to act in ways he believes will bring about that his father leaves him the shirt. It is surely as plausible that he does have this disposition as that he does not. If he believes that he could bring about that his father leaves him the shirt by confessing his situation to his father and asking for his help, or by secretly altering his father's will then it is plausible that he is disposed to act in these ways. And we can easily imagine situations in which he manifests these dispositions and does act in these ways, say if he learns that his father is intent on leaving the shirt to charity or if they have another argument during one of the hospital visits. It seems reasonable to presume that any similar example will be the same: that if it is plausibly a case in which someone has both a self-passive desire that *p* and *I do not act to bring about that p* and a straightforward desire that *p* then it will also be plausible that they have both a disposition to act in

ways that they believe will bring about that *p* and *I do not act to bring about that p* and a disposition to act in ways that they believe will bring about that *p*. So this kind of case is not a counter example to MN.

### 3.8 – Distinguishing Between Desires.

A final objection that I will discuss that is sometimes made to MN on the basis of self-passive desires is that MN cannot distinguish between different desires in respect of their content.<sup>58</sup> It is a condition on an adequate theory of desire that it does at least two things:<sup>59</sup> first, it must correctly distinguish desires from other kinds of mental states. It must say what it is for a mental state to be a desire that *p* as opposed to, for example a belief that *p*, a perception that *p*, an intention that *p*, and so on. Second, it must correctly distinguish between desires that differ in content. It must say what it is for a desire to be a desire that *p* as opposed to a desire that *q*, and so what determines that a desire has the particular object that it does.

According to this objection, MN cannot meet this second requirement of a theory of desire. The problem is supposed to arise in the following way. Recall the earlier example of Mrs. Cunningham: she has a self-passive desire

---

<sup>58</sup> Thanks to Daniel Friedrich for raising this objection to me in discussions.

that Joanie passes her high-school exams on her own, and believes that by encouraging Joanie to study she can bring this about. According to MN Mrs. Cunningham is disposed to encourage Joanie to study. Now consider Mrs. Cunningham\*: she has a desire that Joanie studies, and believes that by encouraging Joanie to study she can bring this about. According to MN Mrs. Cunningham\* is disposed to encourage Joanie to study. So according to MN if they have their respective desires then Mrs. Cunningham and Mrs. Cunningham\* have exactly the same dispositions. Yet, ex hypothesis they have different desires: a desire that Joanie passes her high school exams on her own is not the same as a desire that Joanie studies. So MN does not correctly distinguish between desires that differ in content, so does not meet the second requirement of a theory of desire.

It should be obvious why this objection is misguided. It is an interesting question how to distinguish between desires that differ in content, and we can grant that an adequate theory of desire, a specification of necessary and sufficient conditions for having a desire must provide a way to do this. But MN is not such a theory: it is a necessary condition for having a desire, not necessary and sufficient conditions. So it is not a problem for MN if it cannot distinguish between desires that differ in content in all cases.<sup>60</sup>

---

<sup>59</sup> See, for example (Smith 1998, p449) and (Strawson 1998, p481) for discussion.

<sup>60</sup> Moreover, it is not obvious that MN cannot do this anyway. Consider again the cases of Mrs. Cunningham and Mrs. Cunningham\*. These cases have been mis-described in the objection. It claims that MN entails that they have the same disposition if they have their

### 3.9 – Necessity and Impossibility Desires.

There are two other particular kinds of desire that are said to be counter examples to MN that I will discuss in this chapter. These are what I earlier called ‘necessity desires’ and ‘impossibility desires’. Recall, necessity desires are desires like a desire that there are infinitely many prime numbers, or a desire that  $2 + 2 = 4$ . The propositional object of these desires is necessarily true. Impossibility desires, in contrast are desires like a desire that I have a square circle or a desire that  $2 + 2 = 5$ . The propositional object of these desires is necessarily false. But it is important that in both of these kinds of cases the fact that their propositional objects have their respective truth-values necessarily is something that the desirer is aware of. Distinguish them from

---

different respective desires. But this is false: MN may entail that the dispositions they have are extensionally equivalent in terms of the bodily movements that each is disposed to perform. But they are not the same dispositions. Mrs. Cunningham’s disposition is a disposition to act in ways she believes will bring about that Joanie passes her high school exams on her own. Mrs. Cunningham\*’s disposition is a disposition to act in ways she believes will bring about that Joanie studies. Clearly these are not the same disposition. To put this response a different way: the conditions in which each will be disposed to perform those particular movements are not the same. According to MN, Mrs. Cunningham is disposed to encourage Joanie to study only if she also believes that by encouraging Joanie to study she can bring about that Joanie passes her high school exams on her own. But Mrs. Cunningham\* may be disposed to encourage Joanie to study even if she lacks such a belief. The conditions in which MN entails that they will each manifest their respective dispositions and actually encourage Joanie to study are not the same, so it does not attribute them with the same dispositions. So MN does not entail that having these different desires entails having the same disposition to act. While the objection might raise a problem for a purely behaviourist understanding of having a disposition to act it is not a problem for MN.

cases in which someone has a desire that there are an infinitely many prime numbers or a desire that they have a square circle and does not believe that what they desire is necessarily the case or cannot be the case respectively:<sup>61</sup> MN can account for these other cases in the same way as ordinary desires with contingent propositional objects.

Why are necessity desires and impossibility desires a problem for MN? According to MN, if someone desires that  $p$  then they are disposed to act in ways they believe will bring about  $p$ . So, if someone believes that by doing an action  $m$  they can bring about  $p$  then they are disposed to do  $m$ . But if they have a necessity desire then the object of their desire is something that they believe is already the case, and is the case necessarily. Conversely if they have an impossibility desire then the object of their desire is something that they believe cannot be the case. So it is not plausible that they can also believe that there is some action that they can perform that will bring about what they desire in cases like these. Furthermore, these cases are unlike those where someone has a desire for something that is, or that they believe is contingently possible yet lack a belief about some action they can perform to bring it about. For example, someone might desire that they have a bar of chocolate but lack any belief about something they can do to get one. In that kind of case it is

---

<sup>61</sup> As Schroeder discusses (Schroeder 2004, p16-17), that alternative case is not a counter example to MN: if someone does not believe that what they desire is necessarily the case or necessarily not the case then, just as with other kinds of desire, if they believe that by doing

nonetheless plausible that they are disposed to act as they might easily come to have such a belief about how they can get a bar of chocolate, and if they do have such a belief, and certain other conditions obtain then they will act. For instance, perhaps there is actually chocolate in their fridge and they have merely forgotten that they put it there earlier. They could easily come to remember this, so form a belief that by going to the fridge they can get a bar of chocolate. So it is plausible to attribute them with a disposition to act in this way. But with necessity desires and impossibility desires it is not possible for someone to come to have a belief about some action they could perform to bring about what they desire. Suppose that someone has a necessity desire like a desire that there are infinitely many prime numbers. They are aware that this is the case and is so necessarily. So it is plausible that they cannot also believe of some action  $m$  that by doing  $m$  they can bring about that there are infinitely many prime numbers. They would effectively then believe that by doing  $m$  they could bring about something that is necessarily already the case. But it is not plausible that someone can believe this, or at least, they cannot believe this insofar as they are rational.<sup>62</sup> If it is not possible that someone who desires that there are infinitely many prime numbers can believe of some action  $m$  that by

---

some action,  $m$ , they can bring about what they desire then according to MN they are disposed to do  $m$ . It just happens that in this kind of case this belief will always be false.

<sup>62</sup> So perhaps the explicit target of these cases is MN(rat) rather than MN. But as noted above, MN(rat) is a more restricted thesis than MN so it is entailed by it: so necessity desires and impossibility desire are putatively counter example to MN as well.

doing  $m$  they can bring about that there are infinitely many prime numbers then they cannot have the kind of disposition that MN entails they will have: they cannot have a disposition to act in a way they believe will bring about that there are infinitely many prime numbers. So this is a counter example to MN: it is a case in which someone has a desire that  $p$  and does not have a disposition to act in ways they believe will bring about  $p$ . The same problem for MN will arise, *mutatis mutandis*, if someone has an impossibility desire. So like necessity desires they are counter examples to MN.

### **3.10 – Options (part 3).**

There are a number of responses that are possible here in defense of MN. An initial response is to deny that this kind of case is possible at all. That is, deny that it is possible to have desires for things that you recognize are necessarily the case or that you recognize are impossible. According to this response there are no necessity desires or impossibility desires so there is no counter example to MN here. If someone makes this kind of response they might also claim that those cases where it appears that someone has a necessity desire or impossibility desire have been mis-described. Rather the cases are in fact more accurately described in one of the following ways: (i) it is a case in

which the person is not aware that what they desire is necessary or impossible, so, by definition their desire is not a necessity desire or an impossibility desire. (ii) Or, the person is not in fact rational, so they could believe that there is some action they can perform that will bring about what they desire and they do have a disposition to act when they have their desire. (iii) Or, the kind of mental state they have is not a desire at all. Rather what they have is a hope or wish, or some other kind of pro-attitude towards something that they recognize is necessary or impossible. And as MN is a thesis about desires and not about other kinds of pro-attitude these cases are not relevant to MN.

I think that this response is quite plausible here. In contrast to self-passive desires, actual cases in which someone seems to have a necessity desire or an impossibility desire are not common or familiar.<sup>63</sup> And those actual cases in which someone does claim to have such a desire are not obviously ones where we think the person has fully thought through what it is they desire, or where we think that they are being rational and their desire is a reasonable one for them to have. If there are other cases where the person is obviously rational and properly understands that what they claim to desire is necessary or impossible it is may be plausible that they may have mistaken their attitude for an attitude of hoping or wishing, for example. We are not

---

<sup>63</sup> Although he claims that these desires are possible Schroeder, for example, does acknowledge that they are uncommon. He says, '[d]esires regarding the necessary facts of the world, such as the facts of mathematics, are admittedly scarce, but not as scarce as might be imagined' (Schroeder 2004, p16).



incorrigible about the content of our mental states, so it is plausible that we are not incorrigible about the kind of mental state we are having either.

Alternatively, we can consider this problem for MN as merely an example of a more general problem, what Stalnaker calls the problem of equivalence (Stalnaker 1984, pp. 24-25). This is the problem of how we can stand in different representational relationships to necessarily equivalent propositions. As he says when discussing the problem in the case of beliefs,

the alleged counter examples are not just counter examples to a particular analysis, but cases which are problematic in themselves. We lack a satisfactory understanding, from any point of view, of what it is to believe that *P* while disbelieving that *Q* where the '*P*' and the '*Q*' stand for necessarily equivalent expressions (Stalnaker 1984, p24).

The problem arises from the standard account of propositions, which takes a proposition to be a set of possible worlds at which the proposition is true.<sup>64</sup> But a proposition that is necessarily true is true at all possible worlds, and conversely a proposition that is necessarily false is true at no worlds. So every necessary proposition corresponds to the same set, that is, the set of all worlds.

---

<sup>64</sup> See, for example Lewis (1986): '[A] proposition is a set of possible worlds. A proposition is said to hold at a world, or to be true at a world. The proposition is the same thing as the property of being a world where that proposition holds; and that is the same thing as the set

Analogously, every proposition that is necessarily false corresponds to the same set, but in this case it is the empty set.<sup>65</sup> This leads to a number of familiar problems. For example it seems to rule out the possibility of non-trivial belief in necessary truths: the propositional object of every contingent belief corresponds to a set of worlds that is entailed by the set of all worlds. So every contingent belief entails belief in necessary truths. Moreover, it entails belief in every necessary truth, as all necessary truths are equivalent. Yet intuitively someone can have a contingent belief like a belief that grass is green yet not believe something that is necessary, like a complex mathematical fact: when someone first comes to believe that there are infinitely many prime numbers they gain information that they appeared to be ignorant of beforehand. And what they learn appears to be different from the information they gain when they learn that  $\pi$  is an irrational number. Yet according to this account of propositions if they already had any beliefs about some contingent fact then they already knew both of these necessary facts. So there is a general problem of understanding how we represent propositions that have their truth-value necessarily.

---

of worlds where that proposition holds.' (Lewis 1986, pp. 53-54.) See also Stalnaker (1984).

<sup>65</sup> See, for example Lewis again: 'As possibility amounts to existential quantification over the worlds, with restricting modifiers inside the quantifiers, so necessity amounts to universal quantification. Necessarily all swans are birds iff, for any world  $W$ , quantifying only over parts of  $W$ , all swans are birds. More simply: iff all swans, no matter what world they are part of, are birds. The other modalities follow suit. What is impossible is the case at no worlds; what is contingent is the case at some but not at others.' (Lewis 1986, p7.)

It is plausible that the apparent problems for MN from necessity desires and impossibility desires are just particular examples of this general problem: according to this account of propositions all necessity desires have the same propositional object and all impossibility desires have the same propositional object. But we should not expect to have to solve this general problem in this context. Furthermore, as it is obvious that we can represent such propositions there must be a solution to the problem [see e.g. Stalnaker (1984), especially Ch. 4 and 5 for discussion of suggestions], and it is reasonable to assume that whatever this solution is it will apply to cases of necessity desires and impossibility desires. That is, when we have such a solution we will properly understand what the content of these desires is. It is plausible that when they are correctly understood they will no longer seem to be counter examples to MN.<sup>66</sup>

---

<sup>66</sup> Why is it reasonable to assume this? Note that necessity desires and impossibility desires are a problem for any plausible account of desire. The problem arises for MN because it says that a desire that  $p$  is connected with a disposition to act through having a belief that by doing some action  $m$  I can bring about  $p$ : it is connected with a representation of the object of the desire. And where the object of desire is something that is necessarily the case or impossible there is a general problem that it is not well understood how such things are represented. But any plausible account of desire says that when someone has a desire it is connected in some way with a representation of the object of desire.

Consider, for example, a typical hedonic theory of desire. Hedonic theories attempt to analyse having a desire in terms of having experiences of pleasure. A typical version might claim something like  $S$  has a desire that  $p$  iff  $S$  is disposed to have an experience of pleasure if  $S$  believes that  $p$  has come about [see for discussion (Schroeder 2004, p27)]. Like MN it claims that having a desire is connected in some way with a representation of the object of the desire: in this case, a belief that  $p$  has come about. According to such a version of hedonic theory if someone has any necessity desire then they are disposed to have an experience of pleasure if they form a belief that any contingent state of affairs has come about. Any contingent state of affairs entails everything that is necessarily the case. If someone is rational, is aware that the object of their desire is something that is necessarily the case, and has thought things through then if they believe

### 3.11 – Conclusion.

According to MN, if someone has a desire then, whatever the object of that desire they are disposed to act in ways they believe will bring about that object. But some, like Schroeder for example, have argued that the particular objects of what I have called ‘self-passive desires’, ‘necessity desires’ and ‘impossibility desires’ make them counter-examples to MN. They argue that where someone has one of these kinds of desire then it is not possible that they also have a disposition to act in ways they believe will bring about the object of their desire. So if someone has such a desire then they are not disposed to act as MN entails. However, although these cases appear to be counter-

---

that something contingent has come about they should also believe that anything necessary has come about. So according to this version of hedonic theory having a necessity desire like a desire that there are infinitely many prime numbers disposes someone to have an experience of pleasure if they form any belief that some contingent state of affairs has come about. The converse happens with impossibility desires: it is not possible to believe that something that is impossible has come about, insofar as someone is rational. So it is not possible to have a disposition to have an experience of pleasure if you form such a belief (for analogous reasons that it is not possible to have a disposition to act in ways you believe will bring about something that you believe is impossible, as discussed in the main text). So necessity desires and impossibility desire are counter examples to this account of desire as well, and will plausibly be counter examples to any plausible theory.

If this is right, that these cases are counter examples to any plausible theory of desire then they suggest that it is not possible to give an analysis of desire. But this entails either that desires are *sui generis*, they cannot be analysed in terms of a relation to anything else, or that we do not have desires at all. But neither primitivism about desires nor eliminativism about desires is plausible. It is more plausible that there is something wrong with the supposed counter examples. So it is reasonable to assume that the solution to the general problem can be given to explain away these apparently problematic cases.

examples this is because people have misunderstood what the content of these desires actually is. I have argued that once it is made clear what the objects of these desires are then we can see that they are not counter examples at all. In none of these cases is it plausible that someone has a desire but lacks the disposition to act that MN entails they will have. Indeed, in the case of self-passive desires in particular once the content of those desires is properly understood the original cases that were supposedly counter examples more plausibly provide support for MN instead.

# CHAPTER 4

## DEFENDING DISPOSITIONAL DIRECTION OF FIT

### 4.1 – Introduction.

In Chapter One I highlighted the explanatory power that MN has with respect to some prominent facts about desire. In particular I appealed to the regular correlation that we see between having a desire and acting in ways you believe will bring about what you desire. I argued that this fact is most easily explained if MN is true and I took this as providing support for the thesis. But does this fact provide reason to think there is a connection between desiring and being disposed to act, or is it just reason to think there is a connection between having certain complexes of mental states and being disposed to act? The latter would not support MN, that there is a necessary connection between having a desire in particular and being disposed to act in ways that you believe will bring about what you desire. So why accept my stronger claim?

In fact there is independent reason to think that there is a connection between having a desire in particular and being disposed to act: it provides the

most plausible, informative way of distinguishing between beliefs and desires. The standard way of making this distinction is in terms of what is sometimes called the different ‘Direction-of-Fit’ (hereafter DOF) of each kind of state. But what is it to have a particular DOF, and how does the DOF of belief differ from that of desire. I will argue that having a particular DOF is most plausibly understood in terms of having a particular functional role, that is having a particular set of properties and relations.<sup>67</sup> Furthermore, I will argue that the most informative versions of functional role DOF account are ones that entail that MN is true. Not only does this provide independent support for MN, but it also justifies my earlier claim that the regular correlation between having a desire and acting in ways you believe will bring about what you desire is good reason to accept MN.

## **4.2 – The Explanatory Power of MN.**

Recall my argument in Chapter One: I described a prominent fact about desire, that there is a regular correlation between having a desire and acting in ways you believe will bring about what you desire. It is reasonable to expect that an adequate theory of desire, a set of necessary and sufficient conditions for

---

<sup>67</sup> See Chapter One for a more detailed discussion of what having a particular functional role consists in.

having a desire can explain this fact. I argued that motivational theories of desire, those theories that entail that MN is true, can explain this fact more easily than the standard alternative kinds of theory of desire. I took this to provide support for MN. But is this convincing: what exactly is in need of explanation and does MN have the privileged position for explaining it that I claimed?

It is uncontroversial that there is a correlation between desiring and acting. However, has it been under-described here? After all, it is not the case that there is a regular correlation between someone having just a desire, and no other mental states, and their acting in ways they believe will bring about what they desire. So is the correlation more accurately described as a correlation between having particular complexes of mental states, for example desires and beliefs, and acting in ways you believe will bring about what you desire? If so, then this fact does not support MN to the extent I claimed: according to MN if *S* desires that *p* then *S* is disposed to act in ways that *S* believes will bring about *p*. So MN explains the fact in terms of a necessary connection between having desires in particular, and having a disposition to act. But if the fact that needs explaining is that there is a regular correlation between having certain complexes of, say desires and beliefs and acting then all that is needed to explain it is that there is a necessary connection between having such particular complexes of desires and beliefs and being disposed to act: there



need not be any connection between having a desire in particular and having a disposition to act. So MN does not have a particularly privileged position with respect to explaining this fact. Strawson puts the point in the following way:

Any desire has the following property: it is necessarily true that there are beliefs with which the desire can combine in such a way as to give rise to, or constitute, a disposition to act or behave in some way. This is a conceptual truth, true even of desires to change the past and desires for logically impossible things (Strawson 1994, p276).

But Strawson says this in the midst of a sustained argument against MN. So obviously he does not accept that there is a necessary connection between having a desire in particular and being disposed to act in ways you believe will bring about what you desire. Rather, he thinks that there is a necessary connection between having a particular complex of desires and beliefs and being disposed to act in ways you believe will bring about what you desire. As this is sufficient to explain the regular correlation described above then according to Strawson, that correlation does not provide reason to think that the stronger claim made by MN is true. If Strawson is right then this undercuts what I took to be one of the strongest motivations for accepting MN.

However, as I will argue below there is independent reason to think that the stronger claim made by MN is necessary for explaining this fact. According to what I will argue is the most plausible and informative way of distinguishing belief from desire, it is desire in particular, and not belief, that is connected with having a disposition to act in ways that you believe will bring about what you desire. This entails that MN is true. And it suggests that the stronger claim of MN is justified by this regular correlation.

#### 4.3 – Direction of Fit.

Intuitively a belief that there is chocolate in the cupboard is a different kind of mental state from a desire that there is chocolate in the cupboard.<sup>68</sup> In general we can easily distinguish our beliefs from our desires: we do not usually mistake one kind of state for the other when we have them. But what does the difference between these kinds of state consist in? It is difficult to specify what this is exactly. But the standard current approach is to distinguish between belief and desire in terms of their difference in DOF.

---

<sup>68</sup> Of course, if someone held what Lewis called a ‘Desire as Belief’ (DAB) account of desire then they would deny this [see Lewis (1988)]. According to a DAB account having a desire is constituted by having an appropriate evaluative belief: for example, according to some versions of DAB account *S* desires that *p* iff *S* believes that *p* is good. However, there are good reasons to think that DAB accounts are false (see e.g. Lewis (1988), although see Hajek and Pettit (2004) for further discussion).

According to DOF accounts belief and desire are distinguished from one another by their difference in DOF. Such accounts distinguish two properties: a belief direction of fit (call this 'BDF') and a desire direction of fit (call this 'DDF'). They then claim that being a belief and not a desire consists in having BDF and not having DDF while being a desire and not a belief consists in having DDF and not having BDF. Note that they are not saying that having BDF and not having DDF is sufficient for being a belief, for example. Rather they are saying that it is necessary for being a belief that a state has BDF and does not have DDF. And they are saying that it is sufficient for not being a desire that a state has BDF and does not have DDF. They also claim that being a desire, having DDF and having BDF are related in an analogous way, *mutatis mutandis*. As I understand it the particular aim of DOF accounts is to specify the difference between beliefs and desires, not to give necessary and sufficient conditions for having each kind of state.<sup>69</sup> So all DOF accounts are committed to the following two theses:

---

<sup>69</sup> See e.g. Humberstone (1992) and his objection below. So on this understanding, for example a particular DOF account need not claim that having BDF and not having DDF is sufficient for individuating beliefs from all other kinds of mental states. They might concede that, say visual perceptions also have BDF and lack DDF. But they are committed to saying that having those particular DOF properties is sufficient to distinguish a belief from a desire in particular. Of course, someone could object to this way of understanding the aim of giving a DOF account. However, for the purposes of this dissertation I will follow it: it is adequate for my argument in this chapter in particular.

(T1) For all mental states  $x$  if  $x$  is a belief that  $p$  then  $x$  has BDF and  $x$  does not have DDF.

(T2) For all mental states  $x$  if  $x$  is a desire that  $p$  then  $x$  has DDF and  $x$  does not have BDF.

Of course, merely giving (T1) and (T2) is not yet saying very much of interest. So in addition DOF accounts must say what the respective DOF properties BDF and DDF are. This might be done in various ways by different versions of DOF account. For example, a simple suggestion is to analyse BDF as the property of *being a belief and not a desire*, and to analyse DDF as the property of *being a desire and not a belief*. But obviously such a simple account is unhelpfully circular: it just says that beliefs are not desires and vice versa. But if we are trying to characterise the distinction between these kinds of state then we already know that there is a distinction between them. So this simple DOF account is uninformative. A second version of DOF account is suggested by what Hume says about beliefs and desires.<sup>70</sup> According to Hume beliefs provide information about how the world is whereas desires do not (Hume 1975, p416). Although he did not use these terms, he might be understood as saying that beliefs purport to represent how the world is. In contrast he

---

<sup>70</sup> Indeed, Hume is sometimes credited with introducing this general approach to distinguishing beliefs from desires, although not the name, of course [see e.g. (Sobel & Copp 2001, p44)].

thought that desires did not purport to represent anything at all. He described desires as ‘motivating passions’, he took them to be feelings, like pains or pleasures, except that desires compel us to act in certain ways. Someone might follow what Hume says here in giving a DOF account and characterise BDF as the property of being representational and DDF as the property of being non-representational. Unfortunately for this kind of ‘Humean’ DOF account this is obviously false: the common consensus is that both beliefs and desires purport to represent something. As I discussed in the Introduction a desire is commonly taken to be a particular kind of attitude towards a proposition, and propositions are usually understood as representations in some sense. In this respect they are representational in an analogous way as beliefs are: it is just that believing that there is chocolate in the cupboard and desiring that there is chocolate in the cupboard involve having different attitudes to the same proposition. So this version of DOF account is implausible: it does not even draw the distinction between beliefs and desires correctly so cannot be characterizing the distinction in the right way. A third proposal for characterizing BDF and DDF, suggested more recently by Platts (1979) among others, is in terms of the conditions under which beliefs and desires are in error. According to Platts,

[b]eliefs aim at the true, and their being true is their fitting the world; falsity is a decisive failing in a belief, and false beliefs should be discarded; beliefs should be changed to fit the world, not vice versa. Desires aim at realization, and their realization is the world fitting with them; the fact that the indicative content of a desire is not realized in the world is not yet a failing *in the desire*, and not yet any reason to discard the desire; the world, crudely should be changed to fit with our desires, not vice versa (Platts 1979. pp 256-257).

If we expressed this in the terms of a DOF account it says that a mental state has BDF iff being false 'is a failing' in that state, and a mental state has DDF iff being false is 'not yet a failing' in that state. But again, this is unhelpful. What it is for a particular kind of mental state to have a failing in the sense that is relevant here is not something that is well understood. So if a version of DOF account attempts to characterise the difference between beliefs and desires in terms of something that itself stands in need of analysis then the account is merely postponing giving an informative answer to the question of what distinguishes beliefs from desires.

A more promising proposal is to characterise BDF and DDF in terms of the respective functional roles of belief and desire. Having a particular functional role is a matter of having a particular set of properties and relations:

in the case of mental states these are typically relations with other mental states, sensory stimuli, bodily movements, and so on. Call a version like this a ‘functional role DOF account’. Smith, for example gives a functional role DOF account when he says the following:

For the difference between beliefs and desires in terms of directions of fit can be seen to amount to a difference in the functional roles of belief and desire. Very roughly, and simplifying somewhat, it amounts, *inter alia*, to a difference in the counterfactual dependence of a belief that *p* and a desire that *p* on a perception with the content that not *p*: a belief that *p* tends to go out of existence in the presence of a perception with the content that not *p*, whereas a desire that *p* tends to endure, disposing the subject in that state to bring it about that *p* (Smith 1994, p115).

Expressing this in the terms used by a DOF account, Smith is saying that a particular mental state, *x* that *p* has BDF iff it tends to go out of existence in the presence of a perception as of (with the content that) *not-p*. On the other hand, a particular mental state *x* that *p* has DDF iff it tends to persist in the presence of a perception as of *not-p* and disposes the person having the state to

act in a way that aims at bringing about  $p$ . Smith's DOF account can be summarized by the following two theses:

**(T1-SMITH)** For all mental states  $x$  if  $x$  is a belief that  $p$  then ( $x$  tends to go out of existence in the presence of a perception as of *not- $p$* ) and ( $x$  does not tend to persist in the presence of a perception as of *not- $p$*  and dispose the subject to act in a way that aims at bringing about that  $p$ ).

**(T2-SMITH)** For all mental states  $x$  if  $x$  is a desire that  $p$  then ( $x$  tends to persist in the presence of a perception as of *not- $p$*  and disposes the subject to act in a way that aims at bringing about  $p$ ) and ( $x$  does not tend to go out of existence in the presence of a perception as of *not- $p$* ).

Smith's DOF account avoids the problems faced by the different versions of DOF account I described above: unlike the simple account it does not use the terms being analysed, 'belief' and 'desire' to explain how the states differ. So it does not appear to be unhelpfully circular. Unlike what I called the 'Humean' DOF account, it is not based on an obviously false premise: there is



at least a *prima facie* difference between the effect that having a conflicting perception has on one's beliefs and the effect that it has on one's desires that is like Smith describes. Finally, Smith does not invoke a notion that is itself technical or poorly understood, in contrast to the version of DOF account I attributed to Platts. Rather, Smith's DOF account seems to be appealing to certain relations that are familiar from our own psychology. So, at least in comparison with these alternative kinds of DOF account, functional role DOF accounts such as Smith's appear to give the most plausible characterisation of the difference between beliefs and desires.

#### **4.4 – Counter Examples to Smith's DOF account.**

As I have argued, Smith's DOF account is *prima facie* plausible. However, it will not do as it stands, expressed in terms of the theses (T1-SMITH) and (T2-SMITH). I claimed that it was an advantage of Smith's DOF account, for example over the kind of DOF account I attributed to Platts, that it does not appeal to anything that is itself technical or poorly understood in the way it characterizes BDF and DDF. For example, (T1-SMITH) and (T2-SMITH) describe someone's beliefs and desires as tending to go out of existence or tending to persist if they have certain perceptions. But there are problems

here. There are obvious counter examples to the two theses if these claims about tendencies are understood in what seems a natural way, as statistical claims about when someone will maintain or lose their beliefs and desires. Sobel & Copp (2001) describe cases in which someone will most often maintain a belief that  $p$  despite having a perception as of *not-p*: counter examples to (T1-SMITH). And they describe cases in which someone will most often lose a desire that  $p$  if they have a perception as of *not-p*: counter examples to (T2-SMITH).

Consider Sobel & Copp's example of the first kind case, what they call 'stubborn beliefs' (Sobel & Copp 2001, p47).<sup>71</sup> Suppose that Ayrton, who is an experienced driver, is driving on a hot, cloudless day and believes that the road is dry. Now suppose that Ayrton has a perception as of puddles being on the road ahead of him (a common experience when driving on a hot road), a perception as of the road not being dry. But as Ayrton is experienced he knows that this kind of perception is common in these conditions but that it is illusory. It is plausible that he will tend to maintain his belief that the road is dry when he has this perception. So this is a case in which he has a belief that  $p$  that does not tend to go out of existence in the presence of a perception as of

---

<sup>71</sup> Although I have added details to Sobel & Copp's examples, such as naming the individuals, this is merely to help with the presentation of the examples. I do not take any of the additional details to have any significance for what the examples are supposed to show.

*not-p*. According to (T1-SMITH) Ayrton's mental state does not have BDF so is not a belief: so (T1-SMITH) is false.

Now consider Sobel & Copp's example of the second kind of case, what they call a 'fair-weather fan desire' (Sobel & Copp 2001, p48). Suppose that Brenda is a fan of Gridiron who supports the 49ers in particular: she says that she wants the 49ers to do well. Now suppose that she sees that the 49ers are having a bad streak of losses and sit at the foot of the conference, a perception as of the 49ers not doing well. Brenda no longer says that she wants the 49ers to do well, no longer looks for their scores or cares when they lose. Perhaps, having seen that the Raiders are having a good season, she starts saying that she wants the Raiders to do well. It is plausible that Brenda will tend to lose desires like her desire that the 49ers do well when she had the perception as of the 49ers not doing well. So this is a case in which she has a desire that *p* that does not tend to persist in the presence of a perception as of *not-p*. According to (T2-SMITH) Brenda's desire does not have DDF so is not a desire: so (T2-SMITH) is false.

However, these putative counter examples are unconvincing. It is true that the examples refute (T1-SMITH) and (T2-SMITH) if the theses are understood as statistical claims about peoples' beliefs and desires. But the theses were never intended to be taken in this way. Rather, these claims about tendencies of beliefs and desires to go out of existence or persist are intended

to be understood as claims about the dispositions of someone who has a belief or a desire to maintain or lose that state in particular circumstances. Taken in this way, for a mental state  $x$  that  $p$  to tend to go out of existence in the presence of a perception consists in someone who has this kind of state being disposed to cease having  $x$  if they have a perception as of *not-p*. Conversely, for a mental state  $x$  that  $p$  to tend to persist consists in someone who has this kind of state being disposed to continue having  $x$  if they have a perception as of *not-p*.

Understood in this way it is plausible that Sobel & Copp's cases are not counter examples to Smith's DOF account. Consider again Ayrton the experienced driver with his stubborn belief. For this to be a counter example to Smith's DOF account it must be a case in which Ayrton has a belief that  $p$  and is not disposed to cease having that belief when he has a perception as of *not-p*. But Ayrton's case is not like this. Ayrton is experienced at driving, so it is plausible that he also has beliefs about the unreliability of perceptions you have on hot days in which the road appears wet. In virtue of having this belief it is plausible that Ayrton will be disposed to maintain his belief that the road is dry if on a hot day he has a perception as of the road not being dry. And this disposition may out-weigh a disposition he has to cease having his belief that the road is dry when he has that perception. So the case does not show that Ayrton does not have the disposition to lose the belief that the road is dry (and

it is as plausible that he does have it as that he does not): it is not a counter example to Smith's DOF account.

Now re-consider Brenda with her fair-weather fan desire that the 49ers do well. For this to be a counter example to Smith's DOF account it must be a case in which Brenda has a desire that  $p$  and is not disposed to continue having that desire if she has a perception as of *not- $p$* . But Brenda's case is not like this. It is as plausible that she does have this disposition as that she does not. But it is also plausible that she has other desires, in addition to her desire that the 49ers do well, like a desire that she supports a successful team: this is what makes her a fair-weather fan. In virtue of having such further desires it is plausible that she has a disposition to lose her desire that the 49ers do well if she has a perception as of the 49ers not doing well, and that this disposition out-weighs the disposition she has to maintain her desire that the 49ers do well when she has that perception. So the case does not show that Brenda does not have this disposition: again it is not a counter example.

Smith's DOF account is more accurately characterized in terms of the following expanded theses:

(T1-SMITH\*) For all mental states  $x$  if  $x$  is a belief that  $p$  then, in the absence of counter-veiling beliefs, ( $x$  tends to go out of existence in the presence of a perception as of *not-*

*p*) and (*x* does not tend to persist in the presence of a perception as of *not-p* and dispose *S* to act in ways that *S* believes will bring about *p*).

**(T2-SMITH\*)** For all mental states *x* if *x* is a desire that *p* then, in the absence of overriding desires, (*x* tends to persist in the presence of a perception as of *not-p* and disposes the subject, *S* to act in ways that *S* believes will bring about *p*) and (*x* does not tend to go out of existence in the presence of a perception as of *not-p*).

According to (T1-SMITH\*) Ayrton's stubborn belief that the road is dry does have BDF and does not have DDF. And according to (T2-SMITH\*) Brenda's fair-weather fan desire does have DDF and does not have BDF. So Sobel & Copp's cases are not counter examples.

Note that according to (T2-SMITH\*) there is a necessary connection between having a desire, a state that has DDF and does not have BDF, and having a disposition to act in ways you believe will bring about what you desire. So Smith's DOF account entails that MN is true. If, as I have argued above in section 4.3, this is the most plausible way of characterizing the

difference between beliefs and desires then this is independent motivation to accept MN.

Note also, that according to (T1-SMITH\*) having a belief, a state that has BDF and does not have DDF, does not entail having a disposition to act in ways you believe will bring about what you desire. Recall the objection I discussed to begin with, in section 4.2: according to Strawson, for example, the regular correlation between having a desire and acting in ways you believe will bring about what you desire does not support MN as I had claimed. This is because the correlation is in fact between having certain complexes of beliefs and desires and acting in ways you believe will bring about what you desire. And this correlation is just as well explained by there being a necessary connection between having such complexes of beliefs and desires and being disposed to act in ways that you believe will bring about what you desire as it is by MN, that attempts to explain it in terms of a connection between desires in particular and such dispositions to act. So according to Strawson MN is not necessary to explain the correlation. But if (T1-SMITH\*) is true then if someone has a disposition to act if they have an appropriate complex of beliefs and desires, it is not in virtue of having the beliefs that they have this disposition. So if they have a disposition to act when they have such complexes of states it must be in virtue of having the desires. MN is, in fact needed to explain the correlation. So if Smith's DOF account is true then we

can assume that MN gives the best explanation of this regular correlation: that fact about our psychology is evidence in support of MN.

Of course Sobel & Copp do not genuinely think that their counter examples to (T1-SMITH) and (T2-SMITH) show that Smith's DOF account is false: they did not mis-interpret him as making statistical claims. As I understand them, their counter examples are intended to show that Smith's DOF account must be understood in terms of something like the theses (T1-SMITH\*) and (T2-SMITH\*) if it is to be at all plausible. But note that in each of these expanded theses the term being analysed appears in the analysandum: (T1-SMITH\*) is an analysis of belief and it uses 'belief' on the right-hand side of the bi-conditional. Analogously (T2-SMITH\*) is an analysis of desire and it uses 'desire' on the right-hand side of the bi-conditional. As we will see in the next sections it is this that is the target of both Sobel & Copp's main objection to Smith, and the objection given by Humberstone: that is, that Smith's DOF account is uninformative.

#### **4.5 – Sobel & Copp's Uninformativeness Objection.**

Both Sobel & Copp and Humberstone argue that Smith's DOF account is uninformative. As noted at the end of the previous section, the account



appeals to the notion being analysed in each of its analysis of belief and its analysis of desire. It is this that draws both objections. Nonetheless, the two objections are making quite different points: whereas Humberstone claims that it is the particular way the terms being analysed, belief and desire appear on the right-hand side of the bi-conditionals (T1-SMITH\*) and (T2-SMITH\*) that is objectionable about Smith's DOF account, Sobel & Copp object to the appearance of those terms there in any way at all. Sobel & Copp argue that simply in virtue of having these terms 'belief' and 'desire' appear in the analyses is sufficient for Smith's DOF account to be uninformative. In the following section I will discuss Humberstone's objection in more detail. But first, in the rest of this section, I will respond to Sobel & Copp.

Sobel & Copp present their uninformativeness objection as an objection to functional role DOF accounts in general: although they present their objections against Smith's DOF account in particular they claim that they generalize to all accounts of this kind. As we have seen, they argued by counter example that Smith's DOF account is implausible unless it is understood in terms of something like the expanded theses (T1-SMITH\*) and (T2-SMITH\*). This is so that Smith's talk of mental states 'tending to go out of existence', and 'tending to persist' is understood in the correct way, in terms of someone's dispositions to lose or maintain those states in certain circumstances. But in theses like (T1-SMITH\*) and (T2-SMITH\*) each

analysans is used in the respective analysandum: these theses are giving analyses of belief and desire respectively yet each uses that term to give their analysis. Sobel & Copp claim that because of this Smith's DOF account is unhelpfully circular: to understand the analysis of belief in (T1-SMITH\*) someone would already have to know what a belief is, and to understand the analysis of desire in (T2-SMITH\*) someone would already have to know what a desire is. So Smith's DOF account 'cannot claim to explicate the real difference between beliefs and desires' (Sobel and Copp 2001, p49). And as this is the aim of DOF accounts the version given by Smith is uninformative. Furthermore, they argue that any functional role DOF account will be like Smith's in this respect: they claim that any version of functional role DOF account that is not refuted by their earlier counter examples will be unhelpfully circular in a similar way.<sup>72</sup> So they claim that all functional role DOF accounts are uninformative.

As I understand them, Sobel & Copp's objection to Smith's DOF account can be summarized as follows:

---

<sup>72</sup> Sobel & Copp consider a number of variations of (T1-SMITH). For example, they consider versions that describe relations between a belief that *p* and a perception that counts as *not-p*, relations between a belief that *p* and a perception that counts as evidence that *not-p*, relations between a belief that *p* and a mere seeming that *not-p*, and so on (Sobel and Copp 2001, p49). They claim that either each version is refuted by the examples of stubborn beliefs or it must be understood in an expanded way that uses the notion of belief. In the latter case it will be circular like (T1-SMITH\*). Although they do not argue for it in the same way they also claim that this will be true for all versions of (T2-SMITH), *mutatis mutandis* with respect to desire.

- 1) Smith's DOF account uses the term 'belief' in the analysandum of its analysis of belief and it uses the term 'desire' in the analysandum of its analysis of desire.
  - 2) If a DOF account uses the term 'belief' in the analysandum of its analysis of belief and it uses the term 'desire' in the analysandum of its analysis of desire then it is circular.
  - 3) If a DOF account is circular then it is uninformative.
- C) So, Smith's DOF account is uninformative.

#### **4.6 – Responses to Sobel & Copp.**

One way to respond here is to deny the further claim that Sobel & Copp make, that their objection generalizes to show that all functional role DOF accounts are uninformative. According to this response Sobel & Copp have not shown that there are no versions of (T1-SMITH) and (T2-SMITH) that are neither refuted by their counter examples nor use the terms 'belief' and 'desire' in their respective analysis. They merely showed that a number of particular versions face this dilemma. However, I will not pursue this response here. Rather, in the rest of this section I will argue that their objection to Smith's DOF account in particular appeals to a different assumption that is false.

Premises 1 and 2 of Sobel & Copp's argument are true. But why accept premise 3? It is true that some versions of DOF account are uninformative in virtue of being circular. Recall in section 4.2 I described a simple version of DOF account, according to which BDF is the property of being a belief and not being a desire and DDF is the property of being a desire and not being a belief. This version of DOF account is uninformative: it is circular in a way that is unhelpful for understanding the notions of belief and desire. But just because circularity can make a particular DOF account uninformative this does mean that circularity entails that a particular DOF account is uninformative. We do not see this with certain other kinds of analysis, of different notions, so it does not follow that it is true for analyses of belief and desire.

Consider for example, analyses of colours. One kind of analysis that is sometimes given of colours is in terms of psychological dispositions. Roughly, such dispositional accounts of colours claim that having a particular colour consists in an object being such as to cause observers viewing the object to have psychological dispositions to have particular colour experiences in certain circumstances. For instance, a typical version of dispositional account of red might say something like,  $x$  is red iff  $x$  has some feature in virtue of which  $x$  appears red to normal observers in standard observational

conditions.<sup>73</sup> This dispositional account of red is circular: it uses the term being analysed, 'red' on the right-hand side of the bi-conditional. However it is not obvious that this entails that the account is uninformative. Moreover, all dispositional accounts of colour will be circular for a similar reason. And many people think that at least some versions of dispositional account of colours would tell us something interesting about colours if they are true [see for example, Maund (2002) for discussion]. In general, merely being circular does not entail that an analysis is uninformative.<sup>74</sup> But Sobel & Copp have given no reason to think that in the particular case of analyses of the difference between beliefs and desires merely being circular does entail being uninformative. So we can reject premise 3 of their argument.

In addition, Sobel and Copp have given no reason to think that the circularity in Smith's DOF account is unhelpfully circular, like the circularity in the simple version of DOF account I described earlier, as opposed to being merely circular. We can concede that the circularity in that simple account is unhelpful and therefore the account is uninformative. Recall, the analysis given by the simple version of DOF account appeals to nothing other than the terms being analysed themselves: it proposes to analyse each of belief and desire in terms of being the kind of state it actually is and not being the other

---

<sup>73</sup> See for example Maund (2002), and McGinn (1983).

<sup>74</sup> Although, for a contrasting view see for example Bealer (1982) (1997).

kind of state.<sup>75</sup> But Smith's DOF account is not circular in a similar way. The analysis given by Smith's DOF account appeals to a number of relations, including relations between beliefs and other beliefs, desires and other desires, relations between perceptual states and each of belief and desires, relations between desires and dispositions to act, and so on. There is no reason to think that the circularity in Smith's DOF account is unhelpful and that the account is therefore uninformative. So, Sobel & Copp's uninformativeness objection fails.

#### **4.7 – Humberstone's Uninformativeness Objection.**

In contrast to Sobel & Copp, Humberstone does not claim that Smith's DOF account is uninformative merely in virtue of being circular. His objection is not to the fact that (T1-SMITH\*) and (T2-SMITH\*) have the terms 'belief' and 'desire' respectively on the right-hand sides of each bi-conditional. As he says, '[t]he worry is not that some would-be analysis of the concept of belief fails in virtue of employing, in disguise, that very concept; for clearly no such analysis was being offered' (Humberstone 1992, p64). Rather, according to

---

<sup>75</sup> Consider what an analogous simple account of the colour red would be like: such an account might say something like  $x$  is red iff  $x$  has the property of being red. Like the simple version of DOF account, this is circular in an unhelpful way, and therefore it is uninformative. But

what Humberstone calls ‘the *mutatis mutandis* objection’ it is in virtue of the particular way in which these terms appear in the two theses that Smith’s DOF account is uninformative (Humberstone 1992, pp. 63-64). And this is related to the particular aim of DOF accounts, what such accounts are supposed to provide information about.

Recall from section 4.3, as I understand it, the aim of giving a DOF account is usually taken to be to characterise the difference between beliefs in particular and desires in particular.<sup>76</sup> The aim of giving such an account is not, for example, to give necessary and sufficient conditions for having a belief or having a desire. Rather, having BDF and not having DDF is supposed to be necessary and sufficient for being a belief rather than a desire, given that something is one or other of these kinds of state. Conversely having DDF and

---

this does not show that, for example a dispositional account of red that is circular is therefore uninformative

<sup>76</sup> Although he does not explicitly state it in this way, this seems to be an assumption that Humberstone is making and that underlies the *mutatis mutandis* objection. He says ‘you cannot informatively characterise a fundamental disanalogy between the ways in which beliefs and desires relate to their objects by contrasting them in a respect itself specified by reference to one of these two ways. It’s as if one were to suggest that there is the following deep asymmetry between men and women as regards sexuality: whereas a heterosexual man will *not* be sexually attracted to males, a heterosexual woman *will* be’ (Humberstone 1992, p64, original emphasis). He takes a DOF account as characterizing a disanalogy or a deep asymmetry between belief and desire. But to characterise a disanalogy between two things is merely to say how they differ. It is not to say, for example what either of those things is, what are necessary and sufficient conditions for being each kind of thing. So it is plausible that Humberstone does make this assumption regarding the aim of DOF accounts.

Smith seems to make the same assumption. He introduces his account by saying ‘the difference between beliefs and desires in terms of directions of fit can be seen to amount to a difference in the functional roles of belief and desire’ (Smith 1994, p115). He then proceeds to describe what he takes the functional roles of belief and desire to be. So it is plausible that Smith takes the aim of DOF accounts to be to characterise the difference between these kinds of state, and that this is what he takes himself to be attempting to do with his DOF account.

not having BDF is supposed to be necessary and sufficient for being a desire rather than a belief, given that something is one or other of these kinds of state. It is specifically the distinction between beliefs in particular and desires in particular that is being analysed by a DOF account, as opposed to, say the difference between beliefs and everything that is not a belief and the difference between desires and everything that is not a desire.<sup>77</sup> In light of having this aim there are certain constraints on what an adequate DOF account must do. Obviously it must be true: it must distinguish between beliefs and desires correctly so that for any state that is either a belief or a desire it correctly says which kind of state it is. But in addition it must make this distinction in a particular way: it must distinguish between belief and desire in terms of the properties and relations of each kind of state that are constitutive of this difference between beliefs and desires.<sup>78</sup> This imposes a constraint on the ways in which an adequate DOF account can characterise BDF and DDF if the account is to be informative, that is if it is to provide information about the kind of difference that DOF accounts are supposed to be concerned with.

---

<sup>77</sup> This aim might be understood in epistemic terms in the following way: if someone knew of two mental states that one was a belief and the other a desire then the correct DOF account should allow them to distinguish which is the belief and which is the desire and should tell them what being the kind of state it is rather than the other consists in.

<sup>78</sup> Humberstone says that Smith's DOF account 'does not meet a plausible requirement of universality' (Humberstone 1992, p63). However, he does not say explicitly what he means by this. But as I understand his following discussion it is something like the second condition on an adequate DOF account that I describe in the main text.



As I understand his *mutatis mutandis* objection, Humberstone claims that Smith's DOF account fails this second condition. Note the problem is not supposed to be that it is false, that it does not draw the distinction between beliefs and desires in the right place. Rather, according to Humberstone it draws the distinction in terms of the wrong properties and relations. So, it fails to characterise the difference between beliefs in particular and desires in particular, and this is because of the particular way it is circular. Therefore, it is uninformative: it does not provide the kind of information that a DOF account is supposed to.

The problem, according to Humberstone, is that both the analysis that Smith's DOF account gives of a belief and the analysis it gives of desire appeal to the notion of belief. This is because both kinds of state are analysed in terms of their relations with a perception as of *not-p* and, according to Humberstone, a perception as of *not-p* is itself a state that has BDF and does not have DDF.<sup>79</sup> In this respect a perception as of *not-p* is like a belief and not like a desire.<sup>80</sup> So both kinds of state are being analysed in terms of their relations with something that is belief-like. So Smith's DOF account is circular: it uses a notion being analysed, 'belief' in its analyses. But it is not

---

<sup>79</sup> For example, he says that 'it is clear that we are here explicating the [belief] direction of fit by reference to states with the [belief] direction' (Humberstone 1992, p63).

<sup>80</sup> But to emphasize, having BDF and not having DDF does not entail that a perception as of *not-p* is a belief. Again, having these DOF properties is not claimed to be sufficient for being a belief. Rather it is necessary for being a belief. And it is sufficient for being a belief rather than a desire given that something is one or the other.

this that Humberstone is objecting to. Rather according to him it is the way that each of belief and desire are said to relate to something that is belief-like, a perception as of *not-p*, that is problematic. Whereas Smith says that if *x* is a belief that *p* then it tends to go out of existence in the presence of a belief-like state as of *not-p*, he says that if *x* is a desire that *p* then it tends to persist in the presence of a belief-like state as of *not-p*. But to ‘tend to persist in the presence of something’ is, according to Humberstone to not ‘tend to go out of existence in the presence of something’. Call the relation of tending to go out of existence in the presence of something relation ‘*R1*’. Then according to Humberstone, Smith’s DOF account says that if *x* is a belief that *p* then it has relation *R1* to a belief-like state as of *not-p*, and if *x* is a desire that *p* it does not have relation *R1* to a belief-like state as of *not-p*. The problem should now be clear. A DOF account is supposed to characterise the difference between beliefs in particular and desires in particular. According to Humberstone, Smith’s DOF account claims that this difference consists in beliefs having relation *R1* to something that is belief-like and desires not having relation *R1* to something that is belief-like. This account might in fact be true: it is plausible that it will correctly distinguish between beliefs and desires. For any mental state that is either a belief or a desire Smith’s DOF account will correctly determine which kind it is. However, it does this in terms of the wrong relations: it makes this distinction in terms of beliefs having a certain

relation to something that is belief-like that desires do not have. But, many kinds of thing are like desire in respect of not having relation *RI* to something that is belief-like. For example, it is plausible that intentions, sensations, and emotions are like desires in being different from belief in this way: that they do not have relation *RI* to something that is belief-like. So, according to Humberstone, Smith's DOF account does not distinguish beliefs and desires in terms of the features of each kind of state that are constitutive of the difference between belief in particular and desire in particular. So, it is uninformative.

Consider an analogy for illustration. Suppose you were trying to characterise the difference between dogs in particular and cats in particular. Like beliefs and desires we have little trouble distinguishing between dogs and cats: for example, it is unlikely that someone will mistakenly bring home a Labrador from the pet-shop if they want a cat, nor bring home a Siamese if they want a dog. But suppose that you wanted to know what the difference consisted in. Now suppose you were offered an account of the difference between dogs and cats that said the following: if  $x$  is a dog then  $x$  eats dog-food, and if  $x$  is a cat then  $x$  does not eat dog-food. It is plausible that this account distinguishes between dogs and cats accurately: of all things that are either a dog or a cat, it is plausible that this account will pick out all and only the dogs as dogs, and it will pick out all and only the cats as cats. However, it is clearly inadequate as a theory of the difference between dogs in particular

and cats in particular. It does not distinguish between dogs and cats in terms of the properties and relations of each kind of animal that are constitutive of their difference from animals of the other kind. Rather it does so in terms of the properties and relations common to lots of things that differ from dogs in this way yet are not cats. For example, emus, cows, and goldfish, among other things are like desires in this respect of not eating dog-food yet they are also not dogs. So this relation of not eating dog-food is not the feature of cats that makes them cats and not dogs in particular. This account of the difference between dogs in particular and cats in particular is uninformative.

Similarly, according to Humberstone Smith's DOF account does not distinguish between belief in particular and desire in particular in an informative way. It characterizes the difference in terms of beliefs having relation *R1* to something belief-like, and desires not having *R1* to something belief-like. But as we have seen, this relation is something that desires have in common with many other mental states that are also not beliefs. So this relation of not having relation *R1* to something belief-like is not the feature of desire that makes it a desire and not a belief in particular.

As I understand it, Humberstone's *mutatis mutandis* objection can be summarized in the following way:

- 1) If a DOF account is informative then it analyses belief in terms of the properties and relations that make it a belief and not a desire in particular, and it analyses desire in terms of the properties and relations that make it a desire and not a belief in particular.
  - 2) Smith's DOF account analyses belief in terms of having relation *R1* to something belief-like and it analyses desire in terms of not having relation *R1* to something belief-like.
  - 3) It is not the case that not having relation *R1* to something belief-like is the relation of desire that makes it a desire and not a belief in particular.
- C) So, Smith's DOF account is uninformative.

#### 4.8 – Response to Humberstone.

How might we respond to Humberstone's *mutatis mutandis* objection in defense of Smith's DOF account? The argument is valid, and premise 3 is very plausible. So we can either deny premise 1 or deny premise 2.

One way to deny premise 1 is by challenging the assumption about the aim of giving a DOF account: that is, deny that the aim of these accounts is to provide information about the difference between belief in particular and

desire in particular. If that assumption is false then it would not follow that a DOF account could only be informative if it provided information about that specific difference. So Smith's DOF account would not be uninformative in virtue of failing to do so, as Humberstone claims. We might question whether that is, in fact the aim that people in general take themselves to have when proposing a DOF account. Or we might question why we are interested in that difference in particular, and question whether it is a reasonable constraint to put on a theory. However, I will not respond in this way here: we can grant Humberstone this assumption as his objection fails nonetheless.<sup>81</sup> In the rest of this section I will argue that premise 2 of Humberstone's objection is false.

According to premise 2, Smith's DOF account attributes belief with a particular relation that it says desire does not have. This is true. However, what is not true is that this is all that Smith's DOF account says to characterise the difference between these kinds of state: it does not say that the fact that belief has this particular relation with a perception as of *not-p* and desire does not have this relation with a perception as of *not-p* is what the difference between beliefs and desires consists in. Premise 2 ignores the details given by Smith's DOF account. The account analyses belief and desire in terms of a

---

<sup>81</sup> Indeed, premise 1 seems plausible. As we saw from the references earlier, it is plausible that people do aim to provide this particular information when giving a DOF account. And there might be good reason to be interested in this particular difference: for example, we might take it to be illuminating about the difference between cognitive and conative states in general. Or we might think it is important to understand how the states differ so as to understand how they are able to interact in causing actions, or in practical reasoning, for example.

number of different relations that each state has, not just the relation *R1* with something belief-like described in premise 2. Consider again the expanded theses (T1-SMITH\*) and (T2-SMITH\*) that summarise the account:

(T1-SMITH\*) For all mental states *x* if *x* is a belief that *p* then, in the absence of counter-veiling beliefs, (*x* tends to go out of existence in the presence of a perception as of *not-p*) and (*x* does not tend to persist in the presence of a perception as of *not-p* and dispose *S* to act in ways that *S* believes will bring about *p*).

(T2-SMITH\*) For all mental states *x* if *x* is a desire that *p* then, in the absence of overriding desires, (*x* tends to persist in the presence of a perception as of *not-p* and disposes the subject, *S* to act in ways that *S* believes will bring about *p*) and (*x* does not tend to go out of existence in the presence of a perception as of *not-p*).

According to (T1-SMITH\*) what makes a mental state a belief that *p* and not a desire that *p* in particular is that it has a certain relation, what I called *R1* with a perception as of *not-p*, and a different relation, call it '*R2*', with counter-

veiling beliefs, and it does not have a further relation, call it '*R3*', with dispositions to act in ways that are believed to bring about *p*. And according to (T2-SMITH\*) what makes a mental state a desire that *p* and not a belief that *p* in particular is that it does not have *R1* with a perception as of *not-p* if it also does not have a different relation, call it '*R4*', with overriding desires, and it does have relation *R3* with dispositions to act in ways that are believed to bring about *p*. Clearly, saying all of this is saying much more than just that the difference between beliefs and desires consists in the fact that belief has and desire does not have a particular relation with certain perceptions.<sup>82</sup> So premise 2 is false and the objection fails.

#### 4.9 – Relevance to MN.

Is there an analogous objection to Smith's DOF account when all of the detail in the account is acknowledged? Such an objection will have premises 2 and 3 appropriately modified in the following way:

---

<sup>82</sup> In fact, it is plausible that Smith's DOF account provides an even more detailed characterisation of the difference between beliefs and desires that is given by these theses. It is plausible that the analyses summarized in (T1-SMITH\*) and (T2-SMITH\*) are elliptical for some longer set of properties and relations that beliefs and desires have respectively. Recall, Smith describes his analysis in the text as rough and simplified (Smith 194, p115). And in a note on the passage he says '[w]hen I talk of states having directions of fit, I therefore have in mind whole packages of dispositions constitutive of desiring and believing' (Smith 1994, p210 n8).



- 1) If a DOF account is informative then it analyses belief in terms of the properties and relations that make it a belief and not a desire in particular, and it analyses desire in terms of the properties and relations that make it a desire and not a belief in particular.
  - 2\*) Smith's DOF account analyses a belief that  $p$  in terms of having relation  $R1$  to something belief-like and relation  $R2$  to something belief-like and not having relation  $R3$  to dispositions to act in ways that are believed to bring about  $p$ , and it analyses a desire that  $p$  in terms of not having relation  $R1$  to something belief-like if it does not have relation  $R4$  to overriding desires and having  $R3$  to dispositions to act in ways that are believed to bring about  $p$ .
  - 3\*) It is not the case that not having relation  $R1$  to something belief-like if it does not have relation  $R4$  to overriding desires and having  $R3$  to dispositions to act in ways that are believed to bring about  $p$  is the set of relations of a desire that  $p$  that makes it a desire and not a belief in particular.
- C) So, Smith's DOF account is uninformative.

But premise 3\* is much less plausible than premise 3 which was obviously true. Once the detail in Smith's DOF account is properly taken into account

then it is plausible that it does characterise the difference between beliefs and desires in terms of the properties and relations of each kind of state that are what makes it the kind of state it is and not the other kind of state in particular.<sup>83</sup> So an analogous objection to the properly detailed version of Smith's DOF account will also be unsuccessful.

Note the relevance for MN. According to this properly detailed version of Smith's DOF account a desire that  $p$  has relation  $R3$  to dispositions to act in ways that are believed to bring about  $p$ . Consequently Smith's DOF account entails that there is a necessary connection between desires that  $p$  and dispositions to act in ways that are believed to bring about  $p$ , so it entails that MN is true. In addition, this properly detailed version claims that a belief that  $p$  does not have  $R3$  to dispositions to act in ways that are believed to bring about  $p$ . So it suggests that the regular correlation between having a desire and acting in ways you believe will bring about what you desire is evidence for MN, as I argued in Chapter One. Moreover, the relevance to this dissertation of DOF accounts of belief and desire, and Smith's DOF account in particular is that it entails MN and suggests that this regular correlation provides support for MN. So a version of DOF account like the one Humberstone attributes to Smith, that does not do these things will be irrelevant here. But this properly

---

<sup>83</sup> This is even more plausible if the details Smith gives explicitly in the text are taken to be elliptical for an even more detailed set of properties and relations that each state has (see previous footnote).

detailed version of Smith's DOF account, that entails MN and allows this regular correlation to provide support for MN, is not vulnerable to Humberstone's objection. So to the extent that Smith's DOF account is relevant to my defense of MN then Humberstone's objection fails. If the objection were successful then the DOF account that it showed was uninformative would be irrelevant to my interests here in the first place.

#### **4.10 – Normative Objections to Smith's DOF Account.**

There is one final point I will note before concluding this chapter. I have defended Smith's DOF account against a number of objections, raised by Sobel & Copp and by Humberstone. But according to some people merely doing this ignores what are the strongest objections to MN. People such as Anscombe (1957), Platts (1979), and Zangwill (1998) among others take direction of fit to be a normative notion. As I understand it, they think that what it means to say that a kind of mental state has a particular DOF is that the mental state has certain correctness conditions.<sup>84</sup> So they take the difference between beliefs and desires to be a normative difference, a difference in the

---

<sup>84</sup> Recall the passage from Platts quoted in section 4.3 above. See also Anscombe's example of the shopping list and when it is mistaken that is commonly taken as a way of characterizing direction of fit (Anscombe 1957, p56).

correctness conditions of each kind of state. But, according to people like Zangwill, Smith's DOF account cannot account for any such normative differences between beliefs and desires, so it does not give an adequate account of what the difference between beliefs and desires consists in.<sup>85</sup> According to Smith's DOF account, having a particular DOF consists in having a particular functional role, that is a particular set of properties and relations. And in this case these are natural properties and relations, like a disposition to act, dispositions to lose or maintain a mental state, and so on. But according to the objection natural properties like these do not entail normative properties.<sup>86</sup> So Smith's DOF account does not characterise the difference between beliefs and desire as a normative difference. While beliefs and desires might very well differ in the way Smith claims, according to the objection this difference is not the difference in their respective directions of fit: so it is not what is constitutive of the distinction between beliefs and desire.

I mention this objection to set it aside. As I discussed in the Introduction, I am not concerned with any normative features of desires in this dissertation. I distinguished MN from what I called MN(rat), that says if *S* desires that *p* then *S* is disposed to act in ways *S* believes will bring about *p*, to

---

<sup>85</sup> See also Humberstone, who does not pursue this objection but says, '[w]e note that a further problem with Smith's characterisation [...] is that this normative element is not so much left mysterious as, rather, left out of the account altogether' (Humberstone 1992, p69, n16).

<sup>86</sup> This is like the point familiar from Moore (1903) that 'you cannot derive an "ought" from an "is"', also discussed by Kripke (1982).

the extent that *S* is rational. This objection is targeted more directly to MN(rat) than MN: issues about the correctness conditions of a kind of mental state are issues about when it would be rational for someone to have that kind of state, what it would be rational for them to do if they have that kind of state, and so on. Of course, MN(rat) is entailed by MN, so if this objection is correct then it will refute MN as well as MN(rat). So the objection needs addressing at some point. However, I will not attempt to do so in this dissertation.<sup>87</sup>

#### 4.11 – Conclusion.

DOF accounts attempt to analyse the difference between belief and desire in terms of what they call their different ‘directions-of-fit’. The most plausible DOF accounts characterise this in terms of particular properties and relations that each state has and the other lacks: and according to a plausible version, given by Smith, among the relations that constitute the DOF of desire is a relation with a disposition to act in ways you believe will bring about what you desire. If Smith’s version of DOF account is true this entails that MN is true. In addition, according to Smith’s DOF account it is partly constitutive of the DOF of belief that it does not have such a relation with a disposition to act in

---

<sup>87</sup> I will briefly discuss certain normative features of desire like these in the dissertation’s Conclusion.

ways you believe will bring about what you believe. This is relevant to MN in the context of the explanatory privilege that I argued it has with respect to explaining the regular correlation between desire and acting. If Smith's version of DOF account is true then this correlation is not explained by any properties of beliefs that someone might have when they have a desire and they act. This suggests that it must be explained by properties of their desires: so this correlation provides support for MN. I have defended Smith's DOF account against objections made by Sobel & Copp, and by Humberstone respectively who argued in different ways that it was uninformative. But these objections fail: contra Sobel & Copp, the fact that Smith's DOF account is circular does not entail that it is uninformative. And contra Humberstone, the detail about the different relations given by its analyses of both belief and desire mean that Smith's DOF account can meet the aims of this kind of account: it does provide an analysis that characterizes this difference, between beliefs and desire in particular. So Smith's DOF account gives a plausible characterisation of an intuitive difference, the difference between beliefs and desires: and the account provides further reason to think that MN is true.

# CHAPTER 5

## DESIRE AND REWARD.

### 5.1 – Introduction.

It is often assumed that MN is entailed by a functionalist approach to analysing mental states. However, as we saw in Chapter One, this is a mistake: functionalism is a method for deriving an analysis of a particular thing from a set of claims about properties and relations that are taken to be essential characteristics of that kind of thing. So a particular version of functionalist analysis of desire will only entail MN if it is among the claims that are taken to be essential to desire. If someone thinks that MN is false then they deny that it should be among these initial claims: the objections that I have been addressing in previous chapters are, in effect, attempting to show this. These objections claim that the particular relation described by MN, between desire and dispositions to act, is not a relation that is essential to desire. But a different kind of objection has recently been made by Timothy Schroeder. Schroeder claims that the relation MN describes between desires and

dispositions to act is not even the kind of relation that should be described by an analysis of desire. He thinks that an analysis that posits claims about this kind of relation will not yield an informative account of desire. Rather, it should be describing the kind of relations that are discovered by scientific investigation of desire.

Furthermore, Schroeder argues that the best scientific theories suggest that this relation with dispositions to act is not necessary for having a desire: so MN is not consistent with the discoveries of empirical science. This is what I call his 'Natural Kinds objection'. Instead, according to Schroeder it is a relation with reward that science suggests is necessary for having a desire. So in addition to arguing against MN Schroeder argues in favour of an alternative account, what he calls a 'Reward Theory of Desire'.

Suppose that we grant Schroeder's claim about the kind of relations that desire should be analysed in terms of. Is MN inconsistent with what best scientific theories say about desire? And do these theories suggest that it is a relation with reward that is essential to desire instead, as Schroeder claims? In this chapter I will argue that Schroeder fails to show either of these things. The scientific information that he takes to support his reward theory in fact provides as much, if not more support for MN: the Natural Kinds objection does not show that MN is false. Moreover, I argue that the most plausible versions of reward theory appear to entail that MN is true.



## 5.2 – Two Functionalist Methodologies.

Schroeder's claim about the kind of relations that should be described by an analysis of desire appeals to a distinction commonly made between two approaches to giving a functionalist analysis: this is the distinction between what are sometimes called 'analytic functionalism' and 'empirical functionalism'.<sup>88</sup> One way of understanding this distinction is in terms of their different methodological assumptions. Both agree that an analysis begins from what are taken to be the platitudes and near platitudes, the accepted

---

<sup>88</sup> It is uncontroversial that people do make this distinction, see for example Braddon-Mitchell and Jackson (1996), esp. Ch. 5, for discussion, although people sometime use different terms for each approach: for example, Block uses the term 'psycho-functionalism' for what I am calling 'empirical functionalism', while Braddon-Mitchell and Jackson use 'common-sense functionalism' and 'analytic functionalism' interchangeably. However, it is not exactly clear how to understand this distinction. For example, according to some the approaches differ just in terms of the content of the claims that they derive an analysis from: that is, they differ with respect to the constraints each imposes on what these claims can be about. Whereas analytic functionalists derive their theories from platitudes and near platitudes about the kind of relations familiar from the everyday way of understanding a mental state, empirical functionalists derive their theories from claims about relations and properties of that state discovered by scientific investigation. A different way of understanding the distinction is in terms of the kind of theory that each takes themselves to be giving: whereas analytic functionalist take themselves to be giving a theory of the meaning of the term being analysed, say 'desire', empirical functionalists take themselves to be giving a theory about the actual properties and relations of the state someone is in when they have a desire. They take the claims about relations familiar from the everyday way of understanding desire to merely fix the reference of the term 'desire', and thereby determine which state it is that has to be investigated scientifically to provide a theory of desire. I am not claiming that the way of understanding the distinction that I describe in the main text is the only way of doing so [and see, for example Braddon-Mitchell and Jackson (1996) for further discussion]. However, I take it as a good way to illustrate the point that Schroeder is trying to make, and from personal communication with him it seems plausible that he would agree with this account.

truths, about a particular kind of state, for example concerning its properties and relations. And they agree that the analysis that you end up with must be consistent with at least some of these: there must be some connection between the analysis that is given of a particular state and what we ordinarily take that state to be like. However, they differ with respect to the kind of relations they respectively analyse desire in terms of, they also differ with respect to the extent to which the analysis can be revisionary, and they disagree about what information can justify any such revisions. Analytic functionalists typically think that particular platitudes will only be revised to the extent necessary for the analysis to be internally consistent, and to be consistent with analyses of other kinds of state, for example, that have been formulated in a similar way. They tend to permit only moderate revision of the concept. In contrast, empirical functionalists tend to permit a larger extent of revision if necessary: they allow that an analysis of a particular kind of state can differ to a greater extent from the way in which that state is ordinarily understood. Empirical functionalists typically allow their analyses to deny a greater proportion of the platitudes about a state than analytic functionalists. In addition they think that this can be justified by conflicts between such platitudes and the claims made by empirical sciences, for example, about how the brain works, and so on. In contrast analytic functionalists tend to allow revisions only on the basis of conflict between the widely accepted truths themselves.

These differences can affect the kind of analysis each kind of functionalism tends to give. Consider how they might each proceed when giving an analysis of desire. At first both the analytic and the empirical functionalist will follow the method of giving a functionalist analysis I described in Chapter 1. Recall, this begins by collecting together the platitudes and near platitudes about desire, the claims about particular properties and relations of desire that are widely taken to be true. These might include claims like, that if someone has a desire then they tend to act in ways that they believe will get them what they desire, that people tend to find it pleasurable to get what they desire, and so on. They will attempt to make these claims more precise: in these cases for example they might precisify to yield the claims, if  $S$  desires that  $p$  then  $S$  is disposed to act in ways  $S$  believes will bring about  $p$ , and, if  $S$  desires that  $p$  then  $S$  is disposed to have experiences of pleasure if  $S$  believes that  $p$  has come about. At this stage there might be some revision of, or rejection of some of these claims where they conflict with others that are taken as being about features that are believed to be more important to our understanding of desire. There might also be some similar revision of, or rejection of some claims where they conflict with established facts about, for example other mental states. They will then conjoin all of these precisified platitudes and near platitudes about desire, and rearrange this conjunction to make desire the subject of the sentence: this will

be a sentence something like, 'if  $S$  desires that  $p$  then  $S$  is disposed to act in ways  $S$  believes will bring about  $p$  and  $S$  is disposed to have experiences of pleasure if  $S$  believes that  $p$  has come about'. Next they identify and replace with variables all of the psychological terms in the conjunction, to get an open sentence like, 'if  $S$   $R1$ 's that  $p$  then ( $S$  is disposed to  $R2$  in ways  $S$   $R3$ 's will bring about  $p$ ) and ( $S$  is disposed to have  $F$  if  $S$   $R3$ 's that  $p$  has come about)'. They will then existentially quantify the sentence and present it as necessary and sufficient conditions for having a desire, something like the following:  $S$  desires that  $p$  iff  $[\exists(R1, R2, R3, F)$  If  $S$   $R1$ 's that  $p$  then ( $S$  is disposed to  $R2$  in ways  $S$   $R3$ 's will bring about  $p$ ) and ( $S$  is disposed to have  $F$  if  $S$   $R3$ 's that  $p$ )] and [ $S$   $R1$ 's that  $p$ ].

At this point the analytic functionalist will typically stop and take this sentence as their analysis of desire. This is supposed to be a conceptual analysis of what we mean by the term 'desire'. It says that a desire is a state that has these properties and relations, and that having a desire consists in being in a state that has such properties and relations. But it does not say what kind of state people are in when they have a desire, for example whether this is a physical state, a non-physical state, a state of the brain, and so on, except for what was already claimed about this in the original platitudes that the analysis was derived from.

In contrast the empirical functionalist will not take themselves to be finished in giving their analysis yet. Rather than taking this as a conceptual analysis of what we mean by the term 'desire', they take it as fixing the reference of that term: it says what we are attempting to pick out when we use that term. They then look to science, for example to investigate what kind of state, or states have these properties and relations. Or if nothing has exactly this set of properties and relations they investigate states that are a close fit, that have most of them, or most of the ones that are taken to be most important: they take this to be the state that we are in when we have desires, and what we are, in fact attributing to people when describing them as having a desire.<sup>89</sup> It is sometimes said to be the state that 'realises' desire is us. Suppose that some scientific investigation is carried out and it is found to be a particular state of the brain that has these properties and relations, say a state of D-fibre activation. And suppose that this investigation discovers certain things about the properties of active D-fibres, for example, what neural connections they have with other parts of the brain, what typically causes us to

---

<sup>89</sup> Of course, problems can arise here if it is discovered that there are no states that have enough of these properties and relations to deserve the name 'desire'. And it seems that an empirical functionalist must accept that this is an open possibility. There are different ways to respond to this. For instance, you might take this to show that there are no desires and that our everyday talk about the mind in terms of desires is straightforwardly mistaken. A different response is to take this as a reductio of the original analysis that was taken to fix the reference of 'desire'. If we took the assumption that we do have desires to be something that we are so sure of that any theory that denies this must be false then we might respond in this way. The problem then becomes how to give an analysis of desire that is correct. Other kinds of response are also possible: which is most appropriate can depend on the particular case, the particular thing that the analysis is of.

be in a state of D-fibre activation and what being in this kind of state typically causes, and so on. An empirical functionalist will then attempt to incorporate this information into their analysis of desire. One way they might do this is to revise the analysis they derived from the platitudes about desire, the analysis that the analytic functionalist takes as a definition of desire. For example, they might alter that list of properties and relations that the analytic functionalist claims something must have if it is a desire: adding to that list some of the properties and relations of active D-fibres that were discovered by the previous scientific investigation, and deleting from that list certain properties and relations that active D-fibres do not have. Their subsequent, revised analysis of desire then says that having a desire consists in having something that has the properties and relations that active D-fibres have.

Note that this is still a functionalist analysis of desire despite the fact that it incorporates empirical information about the actual state that we have when we have desires: an empirical functionalist will not say that having a desire consists in being in a state of D-fibre activation, in the way that a type-type identity theory of desire might do. Rather, they say that having a desire consists in being in a state that has the properties and relations that active D-fibres do. So they do not identify desires with the state that is discovered by science to realize desire in us. So they are not straightforwardly vulnerable to an objection of chauvinism and their theory of desire permits that desires can

be multiply realizable (provided that other kinds of state can have the same properties and relations as active D-fibres).<sup>90</sup>

Why do analytic and empirical functionalist differ in this way? As I understand it they disagree about what each considers it necessary for a good analysis to do. Empirical functionalists consider the kind of conceptual analysis that an analytic functionalist will typically give to be uninformative. What this kind of analysis does is make explicit the platitudes and near platitudes that people who understand the concept of desire already know so it does not provide any new information about desire to those who are familiar with desire: and this includes the same people who are trying to give an analysis of desire and so plausibly consider themselves not to know all there is to know about it. In particular this kind of analysis leaves unanswered questions like ‘why do desires have the particular properties and relations the analysis says that they do?’, and ‘why do we have this particular concept of desire, that leads us to produce this analysis, as opposed to having a different concept?’.<sup>91</sup> As I understand them, empirical functionalists make a

---

<sup>90</sup> See for example, Block (1978), and Lewis (1983) for discussion.

<sup>91</sup> Note that the kind of analysis given by an empirical functionalist can give answers to these questions. This kind of analysis will typically claim that having a desire, for example consists in having a state the properties and relations of a physical state of the body or brain. This will be likely to include natural properties and relations, like causal properties and relations. The empirical functionalist can then answer, for example the question ‘why do desires have the particular properties and relations the analysis says that they do?’ by saying that this is because of the causal properties of the kind of natural state that desires are. They might answer the second question I mentioned, ‘why do we have this particular concept of desire, that leads us to produce this analysis?’, by saying that the properties of the state we

methodological assumption that a good analysis of something must provide new information about that thing, even to those who understand the concept.<sup>92</sup> So by finding out which among the competing analyses is true, if any are, they will learn something. I will call this the empirical functionalists' 'information condition': when applied in the context of a particular analysis it might entail that certain further constraints are imposed about what that analysis can be like, depending on background assumptions about the particular thing being analysed and what must be true of it so that it is possible to investigate it scientifically so as to provide the new information required.<sup>93</sup>

It is plausible that Schroeder is an empirical functionalist in these respects. For example, he describes the procedure he will follow to formulate his theory of desire in the following way:

I will begin with a combination of familiar philosophy and everyday common sense, describe scientific findings that either support or challenge what is believed, interpret the findings in the light of previous theory, and conclude with a new view (Schroeder 2004, p6).

---

are in when we have desires determines what this concept is like, and we have this concept in particular because we are most familiar with desires from our own case.

<sup>92</sup> As I understand him, from private email correspondence, Schroeder makes this kind of complaint about the kind of analysis given by an analytic functionalist and makes this methodological assumption that I am calling the 'information condition'.

<sup>93</sup> As we will see in the next section.



If the procedure I have described above, and that Schroeder claims to be following here is an accurate description of the empirical functionalist method then Schroeder is attempting to give an empirical functionalist analysis of desire. In addition, he seems to impose the information condition, or rather a further constraint that is plausibly entailed by it in this case, as a premise in his objection to MN. This is what I call the ‘Natural Kinds objection’, and will discuss in the following sections. It can be understood as a different kind of objection to MN from those I have discussed in earlier chapters: whereas those objections claim that MN conflicts with certain features and relations that are part of our common-sense understanding of desire, the Natural Kinds objection claims that MN conflicts with empirical information about the states that realize desires, that we are in when we have desires.

### **5.3 – An Information Condition on a Theory of Desire.**

As I understand him, Schroeder makes this methodological assumption of empirical functionalism, what I am calling ‘the information condition’, as a premise in his objection to MN: he takes it that an adequate theory of desire should provide us with new information. It should provide us with information about desire that we did not already know just in virtue of our

everyday understanding of desire. He takes this to impose the following constraint on a theory of desire: that whatever properties and relations a particular theory claims are necessary and sufficient for having a desire it must be consistent with desire being a natural kind (Schroeder 2004, p6).<sup>94</sup> By this he means that whatever a particular theory says is involved in having a desire, it must reflect that a desire is ‘a meaningful, unified scientific entity’ (Schroeder 2004, p6). As I understand it, this is so that desire can be the kind of thing that we can get new information about by scientific investigation, for example information about how having a desire causes the phenomena that are characteristically associated with it. According to Schroeder this particular constraint is implicit in both everyday and philosophical discourse about desire.

In fact Schroeder’s claim is quite plausible here. Consider, first, our everyday understanding of desire. We use the term ‘desire’ when making predictions about and giving explanations of peoples’ behaviour, their emotional reactions, their experiences of pleasure and displeasure, and so on. For example, if James says he desires some chocolate we might expect him to go to the cupboard to get himself some, we might expect him to find it pleasurable when he eats it, and we might expect that we could bring a smile

---

<sup>94</sup> Indeed, Schroeder states this constraint explicitly in his introduction: ‘This phenomenon [what is called ‘desiring’, ‘wanting’ or ‘wishing’ and what Schroeder refers to as ‘desiring’] is a distinct, unified entity, a natural psychological kind’ (Schroeder 2004, p5).

to his face by offering to share our own chocolate with him or make him upset by sneaking ahead and hiding the chocolate from him. Conversely, if we saw James doing these things and having these reactions we might attribute him with a desire for chocolate by way of explanation. But according to Schroeder when we do this we are not merely using the term 'desire' to refer to these characteristic phenomena. He says,

[p]eople normally assume that there are things, desires, the having of which robustly explains a myriad of associated phenomena: actions, feelings of joy and resentment, urges, the acquisition of new desires, and so on. (Schroeder 2004, p178.)

That is, we take ourselves 'to contain certain mental states, desires, which explain why we act as we do and feel as we do' (Schroeder 2004, p178). This seems plausible: that in everyday discourse we do understand having a desire to be a matter of having a certain kind of unified entity or, as I will say from now on, being in a certain kind of unified state.

Is this also true for philosophical discourse about desire, in particular in Philosophy of Mind? According to Schroeder, although people rarely state it explicitly, it is a working hypothesis that desire is a psychological natural

kind:<sup>95</sup> this is suggested by the way that people proceed when trying to give a particular theory of desire. Whether people are trying to analyse desire in terms of motivation and actions, experiences of pleasure, or reward, for example, they

recognize that there are a host of psychological phenomena associated with desire, and all three seek to explain these phenomena by first postulating a discreet psychological entity and then conjecturing about the relations between this entity and the various phenomena. And as psychological phenomena, motivation and pleasure are reasonably discrete entities (Schroeder 2004, p177).

In contrast, if someone took the term ‘desire’ to refer just to certain characteristic phenomena then, according to Schroeder all that they would need to do to give an analysis of desire is identify which phenomena people do typically associate with desire and then say that having a desire consists in having any collection of states that someone might have when they have most of these phenomena.<sup>96</sup> But this is not how people generally go about giving a

---

<sup>95</sup> However, distinguish this from a different idea: Schroeder does not say that it is a common assumption that desire is a physiological or biological natural kind (Schroeder 2004, p177). While that might be entailed by particular theories of desire, for instance a type-type identity theory, it is not an assumption that is generally shared.

<sup>96</sup> As an example of what such a theory of desire would be like Schroeder describes what he calls ‘the mix-and-match theory’. This theory claims that “a desire is any mental state

particular theory of desire in Philosophy of Mind, and Schroeder takes this to show that they do make something like an assumption that having a desire is a matter of being in a certain kind of unified state. If this is right, and both our everyday and philosophical understanding of desire assumes that having a desire consists in being in a particular kind of unified state then Schroeder's constraint on an adequate theory of desire appears to be reasonable.

#### **5.4 – The Natural Kinds Objection.**

Suppose we grant Schroeder this particular constraint, imposed by the information condition, that an adequate theory of desire must be consistent with desire being a natural kind. So, whatever the properties and relations that are essential to desire there must be a unified state that someone is in when they have a desire that has these properties and relations. But as we will see, this constraint seems in tension with our best current scientific information about the brain states that people are in when they have these various

---

having most of the features human beings associate with the term 'desire'." (Schroeder 2004, p177.) The mix-and-match theory specifies (for example) three conditions: (a) having a disposition to act, (b) having a disposition to experience pleasure, and (c) having the capacity to represent something as a reward. It claims that it is sufficient for having a desire that a subject has any two of (a), (b), or (c).

phenomena that are characteristically associated with desire. This leads to what I will call Schroeder's 'Natural Kinds objection'.

What are the phenomena that are commonly associated with desire?  
According to Schroeder,

[i]n the case of desire the familiar features surrounding it are complex indeed. People's desires lead them to travel to Cuba, to develop insomnia, to get butterflies in their stomachs; they gain and lose desires by maturing, by painful experience, by massive head injuries; they know what they want, they do not know what they want, they are sure they hate what they actually want very much. But in all this complexity, three phenomena stand out: motivation, pleasure, and reward (Schroeder 2004, pp. 5-6).

He takes the relations that desire has with motivation, pleasure, and rewards as commonly understood as being essential to having a desire. So it is these relations that someone would begin from to derive an analytic functional analysis of desire. Unsurprisingly they are the kind of relations that the standard alternative theories of desire claim are necessary for having a desire. These relations can be specified more precisely, similar to the way that they are described in such theories: for example, we can plausibly understand the

relation between desire and motivation as a relation between having a desire and being disposed to act in ways that you believe will bring about what you desire. We can understand the relation between desire and pleasure as a relation between having a desire and being disposed to have experiences of pleasure if you come to believe that what you desire has come about. And we can understand the relation between desire and reward as a relation between having a desire and what you desire being a reward for you.

In light of the constraint imposed by the information condition, there must be a unified state that someone is in if they have a desire that has these relations with being disposed to act, with being disposed to have experiences of pleasure, and with something being a reward for you.

However, according to Schroeder the results of recent research in neuroscience suggest that in humans there is no unified state that has all of these relations. Rather, each phenomenon is caused by a kind of neurological state that is distinct from the kind of state that causes the others: the areas of the brain responsible for these different phenomena are distinct from one another. Schroeder claims that the best current theories in neuroscience suggest that the areas of the brain responsible for actions are the supplementary motor area (SMA) and anterior cingulate cortex (ACC) that send signals to the motor cortex which then sends outputs to the spinal cord (and the (SMA) and (ACC) also stimulate the spinal cord directly themselves)

(Schroeder 2004, p110). In addition the best neuroscientific theories suggest that the neurological structure responsible for having experiences of pleasure (and displeasure) is the perigenual region of the anterior cingulate cortex (PGAC), found in the region known as Brodmann's area in what is the phylogenetically old area of the cerebral cortex (Schroeder 2004, p76). Finally, such theories suggest that the neurological structures responsible for things being rewards for someone are two immediately adjacent neurological structures deep in the brain, the ventral tegmental area (VTA) and the pars compacta of the substantia nigra (SNpc) (Schroeder 2004, pp. 49-50). These structures are together commonly called the 'reward system'. These different structures are distinct in terms of their location in the brain and also in their relations with the respective phenomena that each is involved in causing. If this neuroscientific information given by Schroeder is true then there is no unified state that humans have that is related in the right way to their dispositions to act, and is related in the right way to their dispositions to have experiences of pleasure, and is related in the right way to things that are rewards for them. Yet if it is a constraint on what desire can be that if someone has a desire then they are in a unified state that has at least these three relations then we are never in a state deserves the name desire. There is no state that we are ever in that has all of the relations that something must have for it to be a desire, according to this constraint. So we do not have desires.



And consequently, all theories of desire are false as they all attempt to analyse what it is that we take ourselves to have when we have desires. But as it turns out, we never have anything that it is appropriate to call a desire.

Consider an analogy for illustration. Suppose that we were trying to give a theory of football managers and that it is an assumption about football managers, that, just like desires if something is a football manager then it is a unified entity, or, more naturally in this case, an individual person. Such a theory has to explain a number of different phenomena including the selection of the team each Saturday, the substitutions that are made during the game, the buying and selling of different players, the interviews that are given to the media, and so on. Suppose that there are three phenomena in particular that are identified as essential to a club having a football manager, on the basis of which clubs are described as having a manager and in the absence of which a club is commonly said to be manager-less: first, having someone who gives instructions from the bench on match-days, second, having someone who negotiates with players in the boardroom, and, third, having someone who conducts training with the players during the week. According to this assumption, if a club has a football manager then there must be an individual at the club who does at least all of these three things. And any adequate theory of football managers must be consistent with their being individuals at football clubs who do all of these things. Now imagine that an investigation is

conducted into all of the football clubs and it discovers that in each case there is no single person at any club who does all of these three things: at each club the different tasks are carried out by three different people each doing one of the tasks exclusively. So the empirical evidence shows that in all the clubs, having these three things done involves having three different people rather than a single person. So according to this assumption none of the clubs ever have a football manager and every theory of football managers, to the extent that it is analysing this concept is false.

This problem that Schroeder raises seems to rule out the possibility of having a desire, and to show that all theories of desire are false. Call this the ‘Natural Kinds objection’: it can be summarized as follows:

- 1) If  $S$  desires that  $p$  then  $S$  is in a unified state,  $x$ , such that  $x$  causes  $S$  to be disposed to act to try to bring about that  $p$ , and  $x$  causes  $S$  to be disposed to experience pleasure if  $S$  comes to believe that  $p$  has come about, and  $x$  causes  $p$  to be a reward for  $S$ . (Information condition.)
- 2) For all humans,  $H$ , if  $H$  is disposed to act to try to bring about that  $p$  then this is caused by  $H$ 's (SMA) and (AC), and if  $H$  is disposed to experience pleasure then this is caused by  $H$ 's (PGAC), and if  $p$  is a

reward for  $H$  then this is caused by  $H$ 's VTA and SNpc ( $H$ 's reward system). (Best theories from neuroscience.)

- 3) For all humans,  $H$ , it is not the case that there is a unified state  $x$  such that  $x$  causes  $S$  to be motivated to act to try to bring about that  $p$ , and  $x$  causes  $S$  to be disposed to experience pleasure if  $S$  comes to believe that  $p$  has come about, and  $x$  causes  $p$  to be a reward for  $S$  (from 2).
- C) For all humans,  $H$ , it is not possible that  $H$  desires that  $p$  (from 1 and 3).

(Corollary) All theories of desire are false.

### 5.5 – Schroeder's Response to the Natural Kinds Objection.

If the Natural Kinds objection is successful then it entails that we never have desires. A natural response here is to take this as a reductio of the argument: this conclusion is so counter-intuitive that it shows that there is a fault in the argument somewhere. Either, one of the premises must be false, or there must be a mistake in the reasoning at some point. Indeed, this is the kind of response that Schroeder gives. Furthermore, he gives a theory of desire

himself, so obviously thinks that there are desires and that they can be analysed. So, what does Schroeder diagnose as the fault in the argument?

As we saw in section 5.3 above, Schroeder accepts the information condition on a theory of desire. So, he accepts that whatever desire is, if someone has a desire then they are in a unified state that is related in the right way to the phenomena that are essential to desire. And he accepts the information, given by what he takes to be the best current theories in neuroscience, about the different areas of the brain and what they are responsible for. Rather, according to Schroeder, the mistake in the Natural Kinds objection as I have summarized it above is in premise 1. His response is to deny this premise, by denying that all of the three relations it describes are essential to desire: that is, he denies that the unified state that someone is in if they have a desire must be related with having a disposition to act, having a disposition to have experiences of pleasure, and having something that is a reward for you. Rather, according to Schroeder what is necessary is that this unified state is related in the right way with having something that is a reward for you.

We can understand Schroeder as responding to the empirical findings about desire by revising the concept of desire, revising the set of relations that the analysis of desire is derived from. The way he does this is to exclude the relations between desire and being disposed to act and between desire and

being disposed to have experiences of pleasure as necessary conditions for having a desire. But why revise it in this particular way? As Schroeder himself acknowledges, of the three relations he initially claimed are commonly taken to be essential to desire the relation between desire and having something that is a reward for you is the least familiar:

It is uncontroversial and familiar that desiring has something to do with being motivated to act and something to do with feeling pleased or displeased on various occasions. If I want my father to be healthy, for instance, I will tend to do certain things (make sure he sees a doctor when sick, say) and I will be open to certain feelings (unhappiness if he has high blood pressure, relief if a tumor proves to be benign, etc). It is less familiar, but also true, that desiring has something to do with reward and punishment. If I want to play with my nephew, say, then I can be rewarded for mending my relationship with my sister by being given opportunities to play with my nephew, and I can be punished for quarrelling with her by being denied such opportunities (Schroeder 2004, p6).

But according to Schroeder there are convincing counter examples that show that the first two relations are not necessary conditions for having a desire. In

the case of a relation between desire and dispositions to act, these are the kind of examples I have discussed in previous chapters. As I have discussed there, if these counter examples are successful, as Schroeder takes them to be, then MN is false. There are analogous kinds of case that he takes to show that a relation between desire and dispositions to have experiences of pleasure is also not necessary for having a desire.<sup>97</sup> In addition he does not seem to think that there is some other relation, different from this original three that is a plausible candidate to be an essential feature of desire. This leaves the relation between desire and something being a reward as the only relation that is essential to desire: he takes this to show that being in a unified state that has the right kind of relation with having something that is a reward for you is what having a desire consists in.

### **5.6 – Schroeder’s Reward Theory of Desire.**

By this process of elimination in response to the Natural Kinds objection, Schroeder reaches the conclusion that having a desire consists in being in a particular kind of unified state that is related in the right way with having something that is a reward for you. He then uses empirical information about

---

<sup>97</sup> See Schroeder (2004), especially Chapters 1 and 3 for extensive discussion.

this relation to give a more detailed analysis. This is what he calls a ‘Reward Theory of Desire’ (RTD), summarized in the following thesis:

**RTD**                      To have a [...] desire that *p* is to [have] the capacity to perceptually or cognitively represent that *p* to constitute *p* as a reward (Schroeder 2004, p131).<sup>98</sup>

We can see that this analysis is not vulnerable to an analogous Natural Kinds objection. Such an objection, appropriately modified so as to target the phenomena that RTD says are essential to desire would be something like the following:

1\*) If *S* desires that *p* then *S* is in a unified state, *x*, that causes *p* to be a reward for *S*. (Information condition.)

---

<sup>98</sup> Schroeder presents this as a restricted thesis: he takes it to be an analysis of ‘positive’ desires alone, and proposes a parallel account of negative desires, or aversions. I will not discuss that account here. In addition, he proposes RTD as an analysis of ‘intrinsic’ desires. By this he means to contrast them with ‘instrumental’ desires. It is not clear what the exact difference between these kinds of desire is, but one way of understanding it is as follows: a desire that *p* is an intrinsic desire if *p* is desired because of some features of *p* itself. For example, if James desires the bar of chocolate in the cupboard because of the taste of the chocolate then James’ desire is intrinsic. In contrast a desire that *p* is an instrumental desire if *p* is desired as a means to achieving something else that is desired intrinsically. For example, if James desires the bar of chocolate in the cupboard because he believes that by eating it he can annoy his sister Louisa and he desires to do that then James’ desire for the chocolate is an instrumental desire. As I will argue below, the fact that his theory is restricted in this second way is a problem for Schroeder, although it is one that he is prepared to accept [see e.g. (Schroeder 2004, p172)].

- 2) For all humans,  $H$ , if  $H$  is motivated to act to try to bring about that  $p$  then this is caused by  $H$ 's (SMA) and (AC), and if  $H$  is disposed to experience pleasure then this is caused by  $H$ 's (PGAC), and if  $p$  is a reward for  $H$  then this is caused by  $H$ 's VTA and SNpc ( $H$ 's reward system). (Best theories from neuroscience.)
- 3\*) For all humans,  $H$ , it is not the case that there is a unified state  $x$  that causes  $p$  to be a reward for  $H$  (from 2).
- C) For all humans,  $H$ , it is not possible that  $H$  desires that  $p$  (from 1\* and 3\*).

(Corollary) RTD is false.

But premise 2 entails that premise 3\* is, in fact, false. So the argument is unsound: RTD is not vulnerable to this appropriately modified version of Natural Kinds objection.

But is RTD even *prima facie* plausible? Recall, Schroeder denied that the relations between desire and dispositions to act, and desire and dispositions to have experiences of pleasure are necessary conditions of desire on the basis of counter examples to those relations. Yet there seem to be analogous counter examples to RTD: it is plausible that there are cases in which someone desires that  $p$  yet does not represent  $p$  so that it is a reward. Unfortunately, it is



not clear exactly what is involved in representing something so that it is a reward. But we can get some idea of this by considering ordinary cases in which we represent something as a reward, like seeing a 'WANTED' poster that says that there is a \$1000 reward for capturing the outlaw, or a case in which a small child believes that if they behave at Grandma's house then their parents will give them a bag of sweets on the journey home, or perhaps a case in which a monkey believes that by pointing to the researcher's cards in an experiment they can get a banana, so takes the banana to be a reward. Obviously it is not necessary for being able to represent something so that it is a reward that someone has the concept of reward. But these ordinary cases seem have in common that someone makes a positive evaluation of something that they believe that they can get only if they fulfill certain explicit conditions. While this is merely an approximation of representing something so that it is a reward I will not attempt to specify it more precisely. It is plausible that there can be cases in which someone desires something that they do not represent like this. Consider, for example, Peter, who desires that his wife and child are in good health. Intuitively, Peter does not represent his wife and child being in good health as a reward. Undoubtedly he will evaluate positively his wife and child being in good health. But it is unlikely that he will believe that this is something that will only come about if he fulfills certain explicit conditions, in the way that someone seeing the wanted poster believes that they can get the

\$1000 reward only if they capture the outlaw. Peter might believe that there are certain things that he can do to try to bring about that his wife and child are in good health: for example, he might believe that he can bring this about by buying family health insurance, or by insisting that they all eat a healthy, balanced diet. But intuitively the way that he understands his wife and child being in good health to be conditional on his doing these things is not the same way that the person seeing the poster understands that getting \$1000 is conditional on capturing the outlaw. Moreover, it does not seem to be a way of representing his wife and child having good health so that that is a reward for him. So this seems to be a case in which Peter has a desire that  $p$  yet does not represent  $p$  so that it is a reward: a counter example to RTD.

Of course, this counter example is not successful. Schroeder does not mean by 'reward' here the ordinary notion of reward familiar from the examples above. Rather he has a technical notion of reward, the notion that is used in behavioural science, and it is a relation between having a desire and representing something so that it is a reward in this sense that he claims is constitutive of desire. Schroeder illustrates this sense of reward with a famous example from behavioural science, B. F. Skinner's (1948) operant conditioning experiments. In these experiments

hungry pigeons [were placed] in a cage that released food every 15 seconds regardless of the pigeon's behaviour. After a number of trials each pigeon had developed distinctive and consistent behaviour patterns [...] Each behaviour had been produced spontaneously by the pigeon and had been (purely by coincidence) immediately followed by food; the coincidence between spontaneous behaviour and food had caused the pigeon to produce the spontaneous behaviour more frequently, which led to another coincidence of behaviour and food, and so on. Each pigeon's behaviour was thus shaped by the pattern of reinforcements it received following its spontaneous behaviour (Schroeder 2004, p44).

In these cases food is a reward for the pigeons in Schroeder's technical sense: it is used to influence the pigeons' behaviour by processes of reinforcement learning. Of course, the kinds of things that can be a reward for someone, the ways in which someone can be influenced by something that is a reward, the capacities that it can have an effect on, and the ways in which something that is a reward can have its effects can vary for different kinds of individual and for different individuals of a particular kind. For example, something that is a reward need not have its effects by brute conditioning like in the experiments with the pigeons, for example. There are more sophisticated reinforcement

learning processes by which rewards can have their effect in different cases. But these cases illustrate the kind of relation that Schroeder is concerned with when saying that something is a reward for you.

So, in the sense of reward that Schroeder is using in RTD,  $p$  is a reward for  $S$  if  $p$  can be used to influence  $S$ 's behaviour, emotional responses, intellectual dispositions, sensory capacities and so on by processes of reinforcement learning. According to RTD having a desire consists in having the capacity to represent something so that it has these kinds of effects on you. And according to Schroeder, for someone to have such a capacity it is not necessary that they can have representations of the thing that is a reward at what is sometimes called a 'personal level': it is not necessary that they can have, say beliefs about it having these effects. Rather, what is necessary is that they have the right kind of neurological connections such that some kind of representation of that thing can initiate specific reinforcement learning processes.<sup>99</sup> To put this more precisely, having the representational capacity necessary and sufficient for having a desire consists in having two capacities and having them connected in the right way. First, the subject must be able to represent  $p$  perceptually or cognitively. Second, they must have the capacity to represent something so that it is a reward so that it initiates processes of reinforcement learning. Third, these two capacities must be connected such

---

<sup>99</sup> Although Schroeder does not seem to say this explicitly, it is plausible that he thinks that quite rudimentary representations can have these effects.

that when the subject perceptually or cognitively represents that  $p$  this causes (or contributes to causing)  $p$  to be represented so that it is a reward.<sup>100</sup>

In light of this, Schroeder's Reward Theory is more accurately summarized in the following way:

**RTD\***            To have a [...] desire that  $p$  is to have the capacity to perceptually or cognitively represent that  $p$  to initiate the production of reward signals that cause a characteristic, mathematically describable change in the connectivities of units that are describable at an appropriately abstract level.

Note that expressing Schroeder's theory in this way, in terms of RTD\* rather than RTD, does not change the fact that it is not vulnerable to an appropriately modified versions of the Natural Kinds objection. According to one of the clauses of premise 2 of the objection, the best current theories in neuroscience suggest that there is unified state that people are in that causes something to be

---

<sup>100</sup> Thanks to Schroeder for clarifying this in his generous personal correspondence with me. He says that this way of characterizing having the capacity to represent something so that it is a reward is necessary to accommodate two intuitions: first that it is possible for someone to have a desire for almost anything that they are able to represent cognitively or perceptually, so there is some connection between the ability to desire something and the ability to perceive it or think about it. Second that there is more to having a desire for something than merely having the capacity to represent it perceptually or cognitively, so just because someone can perceive or think about something this does not entail that they have a desire for it.

a reward for them. This is a state of a structure in the brain that is sometimes called the 'reward system': in humans this is a state of the VTA/SNpc. In fact, according to these neuroscientific theories, the VTA/SNpc works by producing neural and chemical signals that have reinforcement learning effects on other structures in the brain. These include the areas that are responsible for behaviour and action, emotional responses, experiences of pleasure and displeasure, and so on. In fact the VTA/SNpc is thought to have connections to almost every other structure in the brain, and these connections release the neurotransmitter dopamine at their synapses, a chemical that these other structures are known to have receptors for and be sensitive to. The signals from the VTA/SNpc have effects on, for instance the activity levels of different structures, the threshold level of stimulation at which they become active, and so on. So the VTA/SNpc has the extent of influence, and the right kind of effects for its states to be what causes things to be rewards for people in the sense of reward in RTD\*. And it is taken in neuroscience to be a distinct structure, the reward system, and its state to be unified states. So as before, there is a unified state that people have that causes something to be a reward for them. An appropriately modified version of the Natural Kinds objection to RTD\* is unsound: its third premise is false for the same reason as for RTD.

If Schroeder is right, and his response is the most plausible way of avoiding the counter intuitive conclusion of the Natural Kinds objection then RTD\* is true. In which case, the relation between desire and dispositions to act is not essential to desire: it is not necessary that having a desire entails having a disposition to act in ways you believe will bring about what you desire. So MN is false.

### **5.7 – Alternative Responses to the Natural Kinds Objection.**

Is Schroeder's response to the Natural Kinds objection the most plausible? Is the most plausible way of avoiding the conclusion that we do not have desires to revise premise 1 as premise 1\*: that is, to deny that the relations between desire and dispositions to act and between desire and dispositions to have experiences of pleasure are essential to desire, and say instead that being in a unified state that is related in the right way with something being a reward for you is necessary and sufficient for having a desire?

An alternative response to the objection is to deny premise 2 by challenging Schroeder's claims about the different causes of the three phenomena described in premise 1. According to Schroeder this is what is suggested by the best current theories in neuroscience, but someone could

dispute this, as Katz does (Katz 2004). However, I will not pursue this response here.

Instead, note that Schroeder's response is not the only way of denying premise 1: there is a symmetrical response with respect to each of the other two relations. Schroeder revises premise 1 in terms of the relation between desire and something being a reward for you, to get premise 1\*. But in an analogous way someone might revise it in terms of the relation between desire and dispositions to act, to say that the relations between desire and dispositions to experience pleasure and between desire and reward are not essential to desire. This would lead to something like 1\*\*:

(1\*\*) If *S* desires that *p* then *S* is in a unified state, *x*, that causes *S* to be disposed to act to try to bring about that *p*.

Or someone might revise premise 1 in terms of the relation between desire and dispositions to have experiences of pleasure in an analogous way, to yield something like 1\*\*\*:

(1\*\*\*) If *S* desires that *p* then *S* is in a unified state, *x*, that causes *S* to be disposed to experience pleasure if *S* comes to believe that *p* has come about.



Each of these responses is symmetrical and avoids the conclusion of the Natural Kinds objection, that we do not have desires. So why accept that Schroeder's way of denying premise 1 is most plausible? Recall, according to Schroeder there are counter examples to the each of the other two relations. He would take these examples to show that  $1^{**}$  and  $1^{***}$  are false. However, as I noted above in section 5.5, the counter examples to  $1^{**}$  are the kinds of example that I have argued are unsuccessful in previous chapters. If my arguments are successful then there is no symmetry breaker between  $1^*$  and  $1^{**}$  at least. So  $1^*$  has no privilege over  $1^{**}$  as a more plausible response here. And note that  $1^{**}$  entails MN. So, to the extent that the Natural Kinds objection does not show that Schroeder's own theory of desire,  $RTD^*$  is false, it does not show that MN is false.

If the Natural Kinds objection cannot be sound, yet denying premise 1 and revising it as  $1^*$ , which entails  $RTD^*$  and denying premise 1 and revising it as  $1^{**}$ , which entails MN are equally plausible responses to it then Schroeder has not shown that MN is false. But in fact there are reasons to think that the second of these responses, revising as  $1^{**}$  is more plausible, so that the Natural Kinds objection provides support for MN. This is because  $RTD^*$  seems to face a dilemma. On one horn the theory is ill-motivated, there is no reason to think that it is an analysis of what we ordinarily mean by

‘desire’. On the other horn, to the extent that there is reason to think that it is an analysis of desire it is implausible, it has implications that conflict with our ordinary understanding of desire. Consider the first horn of the dilemma. Recall how Schroeder motivated the theory initially: he appealed to a relation between having a desire for something and that thing being a reward for you that he claims is familiar from our everyday understanding of desire. Suppose we grant that there is such a relation that we are familiar with. Perhaps there is evidence of this from the way we use the terms ‘desire’ and ‘reward’. This will be a relation between having a desire and having something as a reward for you in the ordinary sense of reward. That is, the sense of reward in which someone who sees the wanted poster believes that the \$1000 is a reward for capturing the outlaw, or the sense in which the small child thinks of the bag of sweets as a reward for behaving on the visit to Grandma. Yet RTD\* does not say that desire is related to something being a reward in this ordinary sense. Rather, it says that desire is related to something being a reward in a technical sense, in terms of a connection with processes of reinforcement learning. But there is no reason to think that this relation, between having a desire for something and it initiating processes of reinforcement learning is something that we are familiar with from our everyday understanding of desire. There is no linguistic evidence, for example as there is for the relation between desire and reward in the ordinary sense. So there is no reason to think that RTD\* is

an analysis of our ordinary concept of desire: there is no reason to think that whatever has the kind of relation with processes of reinforcement learning that is described in RTD\* is the mental state that we attribute to someone when we say they have a desire.

In fact, Schroeder is aware of this problem and attempts to motivate RTD\* by appealing to the original relation that there is evidence for, the relation between desire and reward in the ordinary sense. According to Schroeder this ordinary sense of reward coincides with the technical sense, the processes of reinforcement learning. He claims that cases that are ordinarily described as rewards are in fact cases where processes of reinforcement learning are taking place. So he claims that this is what we actually refer to when we use the term 'reward' in the everyday sense. But this is implausible. There are counter examples in both directions: there can be cases in which someone would be described as having something as a reward for them in which it is not associated with any processes of reinforcement learning, and there can be cases in which processes of reinforcement learning are taking place but that would not be described as cases where something is a reward for someone. As an example of the first kind of case, consider Mavis, an avid admirer of the Royal family who is unwell in hospital. In addition Mavis intensely dislikes, and is terrified of, being in hospital to such an extent that she finds it unpleasant even to think of herself in hospital and tries to do

everything she can to avoid having to spend any time there. Suppose that Princess Anne is the patron of the hospital that Mavis is staying at and one morning visits the hospital to officially open a new ward. During her visit she walks around some of the other wards and talks with some of the patients at their bedside. Suppose that Mavis is one of the patients that Princess Anne stops to talk with. Mavis considers herself to have been rewarded by this visit. But it is not plausible that Mavis's representations of Princess Anne visiting her in bed in hospital will cause any processes of reinforcement learning in Mavis that could be characteristic of it being a reward. It is not even clear what such effects might be: that is, it is not clear in what ways her thoughts about Princess Anne visiting her in bed in hospital could have effects, by processes of reinforcement learning, on her dispositions to act, her dispositions to have certain emotional responses and experiences of pleasure and displeasure, her intellectual dispositions and sensory capacities that were characteristic of that being a reward for her.<sup>101</sup> So it is plausible that, in this

---

<sup>101</sup> Consider some possible alternatives: it is not plausible that Mavis's thoughts about Princess Anne visiting her bed in hospital could have characteristic effects on her dispositions to act or her dispositions to have experiences of pleasure, for example. If this is a reward for Mavis then it seems that representing it should effect such dispositions so as to make Mavis more disposed to act in ways that will tend to bring it about, and make her more disposed to have experiences of pleasure if it comes about. If representing this had the opposite kind of effect then, according to Schroeder this would be a punishment for Mavis. But any such thoughts represent that Mavis is in hospital, and as we have described the case having thoughts representing this situation makes Mavis feel displeasure and plausibly makes her more disposed to act in ways that she believes will prevent her having to go into hospital. In addition, it does not seem plausible that Mavis's thoughts about Princess Anne visiting her bed in hospital could have characteristic effects on her sensory capacities or intellectual dispositions. There do not seem to be any particular thoughts or perceptions that are characteristically associated with of this kind of event, Princess Anne visiting you in your

case, there is something that is a reward for Mavis in the ordinary sense but not in the technical sense as her representations of that thing do not have characteristic effects by initiating processes of reinforcement learning.

As an example of the second kind of case, consider a simple perceptual discrimination experiment like an experiment in which a subject is briefly shown in succession two images of diagonal lines with different orientations and must choose which is the closest to the horizontal of the two. As they repeat the task their capacity to discriminate the more horizontal line will improve. This is a process in which a sensory capacity is improved through a process of reinforcement learning, initiated and maintained by their visual perceptions of the lines that become less distorted, or less ‘noisy’, as their ability to discriminate improves. But it is not plausible we would describe the lines in the stimulus as a reward for the subject. It might be true that they also have perceptions of, and thoughts about their improving performance on the task, or of praise from the experimenter, and so on: things that we might think of as rewards for them. But it is not these perceptions and thoughts that are responsible for the reinforcement learning effects. So in this kind of experiment the subject has perceptual representations of something, the lines

---

bed in hospital. So, there seems to be no characteristic ways in which such things can be effected by Mavis’s perceptual and cognitive representations of it. So, in this case Mavis’s perceptual and cognitive representations of Princess Anne visiting her bed in hospital do not have characteristic effects by reinforcement learning processes on Mavis’s dispositions to act, her dispositions to have experiences of pleasure, her intellectual dispositions or her sensory capacities, so this is not a reward for her in the technical sense. Nonetheless, it is a reward in the ordinary sense. So these senses of reward are not always co-instantiated.

in the stimulus, that initiate and maintain a process of reinforcement learning that has characteristic effects on a sensory capacity, so this is a reward for them in the technical sense. But what they represent that has these effects, the lines in the stimulus, is not a reward for them in the ordinary sense of reward.

Contrary to Schroeder's claim the ordinary and technical senses of reward are not co-instantiated. So what linguistic evidence there is in support of there being a relation between having a desire for something and it being a reward for you in the ordinary sense of reward does not support the claim in RTD\*, that there is a relation between having a desire for something and it being a reward for you in the technical sense of having characteristic effects on you through processes of reinforcement learning. So RTD\* is unmotivated: there is no reason to think that it is an analysis of desire, the mental state that we attribute to people when we say that they have a desire in everyday discourse.

Of course, Schroeder could insist that we should accept the analysis given by RTD\* despite these problems, perhaps on the grounds that it has greater explanatory power than alternative analyses of desire. But then it seems he is committed to a significant revision of the ordinary concept of desire if he is proposing an analysis in terms of something for which there is no intuitive support. Yet with such an extent of revision it is then as plausible

that Schroeder is in fact giving an analysis of a different kind of mental state, call it 'desire\*' rather than an analysis of desire.<sup>102</sup>

This brings us to the second horn of the dilemma. Suppose that Schroeder can provide some evidence that it is part of our everyday understanding of desire that has a relation with reward in terms of processes of reinforcement learning like in RTD\*: so he can provide independent motivation for RTD\*. Nonetheless, it is not a plausible analysis of desire: it has a number of implications that conflict with our intuitions about desires. First, it implies that it is only possible to have a desire for a fast car, for example if it is already a reward for you: that is, if you already have the sub-personal capacities, connected in the right way to be able to have states that represent fast cars so that they initiate processes of reinforcement learning. That is, if you can have a desire to have a fast car then fast cars must already be a reward for you. Yet this gets the order of explanation wrong. Intuitively you will only be rewarded by being given a fast car if it is something that you have a desire for. Suppose you were an environmental lobbyist, or you are afraid of driving fast, or you already have a collection of fast cars and are

---

<sup>102</sup> Moreover, this response is especially counter-intuitive here. Recall the dialectic: the natural Kinds objection has an unacceptable conclusion, that we never have any desires. In response Schroeder proposes to deny premise 1 of that objection and revise it to be like 1\*. But 1\* entails RTD\*, which is not a plausible analysis of desire. So revising premise 1 of the objection to 1\* does not avoid the conclusion: it does not give an analysis of desire that is not vulnerable to the objection. It merely says that there is some other kind of state, desire\* that has a certain relationship with processes of reinforcement learning. But this is effectively to concede the objection, that there is nothing that deserves the name 'desire'. It does not respond to it.

bored with them. If you were like this then it seems that you will not have a desire for a fast car: so you would not be rewarded by being given one.

Second, RTD\* implies that there is a deep distinction between intrinsic desires and instrumental desires. Recall from above (see footnote 98), a desire that  $p$  is an intrinsic desire if  $p$  is desired because of some features of  $p$  itself, whereas a desire that  $p$  is an instrumental desire if  $p$  is desired as a means to achieving something else that is desired intrinsically. As I noted, RTD\* is restricted: Schroeder intends in only as an analysis of intrinsic desire. But in fact this is not something that he can avoid anyway: RTD\* is not plausible as an analysis of instrumental desire. It is possible to form a new instrumental desire for something very quickly, for example if you suddenly come to recognize that by bringing it about you can bring about something else that you desire intrinsically. Conversely an existing instrumental desire for something can be lost very quickly, say if you suddenly realize that bringing it about will not, as you previously believed, bring about something that you desire intrinsically. According to RTD\* having a desire for something consists in having specific representational capacities and sub-personal, neural connections in place, as discussed above. So it says that forming a new desire is a matter of forming new representational capacities or neural connections, while losing a desire consists in losing specific representational capacities or specific neural connections that you previously had. It is not plausible that



such capacities and connections can be formed and lost at the same rate as instrumental desires can be formed and lost. So having an instrumental desire cannot consist in having the representational capacities and neural connections for representing something so that it is a reward for you. According to RTD\* having an intrinsic desire does consist in having these representational capacities and neural connections. So it implies that intrinsic desires and instrumental desires are genuinely distinct kinds of mental state (perhaps analogous to the way that beliefs and desires are distinct kinds of mental state). Yet, while Schroeder seems prepared to accept this it seems to conflict with how we ordinarily understand desires: intuitively, intrinsic desires and instrumental desires are varieties of the same basic kind of mental state. It seems to be more than a mere terminological accident, for example, that we use the term 'desire' to refer to both kinds of state. In addition, both are thought to be able to cause and justify actions, and both can be considered in the same way when we engage in practical reasoning. So if RTD\* entails that intrinsic desires and instrumental desires are distinct kinds of mental state then it is not plausible as an analysis of desire as it is ordinarily understood.

So Schroeder faces a dilemma: he wants to avoid the conclusion of the Natural Kinds objection, that we never have any desires, and claims that the most plausible response to that objection is to deny premise 1 and modify it to 1\*. This modified claim entails RTD\*. But it is possible to deny premise 1

and modify it in a different way, to 1\*\*, which entails MN is true. Moreover, Schroeder's response, modifying the first premise of the objection to 1\* is not plausible. Either, RTD\* is not giving an analysis of desire at all, as it is ordinarily understood, but rather is an analysis of a different kind of mental state, or RTD\* is giving an analysis of desire and it is false. As Schroeder's response to the Natural Kinds objection entails RTD\* then if RTD\* is not giving an analysis of desire it is not relevant to the conclusion of that objection. On the other hand if RTD\* is giving an analysis of desire, but one which is false, then it is not a plausible response to the Natural Kinds objection: Schroeder's response is either implausible or is not a response at all. Therefore, the alternative response to the Natural Kinds objection, modifying the first premise of the objection as 1\*\* is more plausible than Schroeder's, and it entails that MN is true. So the Natural Kinds objection does not refute MN.

## **5.8 – The Reward Theory of Desire and MN.**

According to Schroeder, as desires are ordinarily understood we take them to have characteristic relations with rewards and to have characteristic relations with motivation, or dispositions to act. Is there also a relation between

rewards and dispositions to act?<sup>103</sup> In this last section of the chapter I will argue that there is such a relation: I will argue that according to the most plausible way of understanding Schroeder's Reward Theory of Desire, RTD\*, if something is a reward for you then you are disposed to act in ways that you believe will bring it about. This would be interesting in itself, but note also the relevance it would have for the discussion so far. I have defended MN against the Natural Kinds objection by arguing that the conclusion of that objection, that we do not have desires, is unacceptable, and that the most plausible response that avoids this also entails that MN is true. But if, as I will argue below, having something as a reward for you entails that you are disposed to act in ways that you believe will bring it about, and RTD\*, that is entailed by Schroeder's preferred response to the Natural Kinds objection also entails that MN is true. Then MN will not be refuted by the Natural Kinds objection whatever the result of the dialectic with Schroeder above about which is the most plausible response to that objection.

Recall how we summarized RTD\*, in terms of the following thesis:

**RTD\***                      To have [a] desire that *p* is to have the capacity to perceptually or cognitively represent that *p* to initiate

---

<sup>103</sup> Recall, in Chapter Two we saw that there are relations between desires and intentions and between intentions and dispositions to act. Might there be analogous relations in this case, between rewards and dispositions to act?

the production of reward signals that cause a characteristic, mathematically describable change in the connectivities of units that are describable at an appropriately abstract level.

As we have seen from the discussion above, according to RTD\* having a desire that  $p$  consists in being constituted such that certain perceptual and cognitive representations of  $p$  initiate processes of reinforcement learning that have characteristic effects in you. But what kind of effects will these be? One version of RTD\* does not commit to there being a certain kind of effect or certain set of kinds of effects that result from the reinforcement learning processes that are entailed by having a desire. According to this version of RTD\* if  $S$  desires that  $p$  then there are reinforcement learning effects in  $S$  that are characteristic of  $S$  having a desire that  $p$ . It says that if  $S$  has this desire then there are some kinds of effect  $E$  that typically occur in  $S$ , but that  $E$  can be any kinds of effect that can be brought about by a process of reinforcement learning. But this is implausible: it implies that when  $S$  has a desire that  $p$  the effects that are characteristic of  $S$  having this desire are (i) only contingently related to the object of the desire,  $p$ , and (ii) only contingently like those that other humans have if they have a desire for the same object,  $p$ . Consider when people have a desire that they have some chocolate. If someone has a desire

for chocolate they will typically be disposed to act to certain ways that they believed will get them chocolate, they will typically have experiences of pleasure if they eat some chocolate, they will typically believe that chocolate tastes nice, and so on. Now consider a particular person, Steve, who has a desire that he has some chocolate. Whenever Steve has this desire then the only effects in Steve are that he is disposed to whistle 'The Blue Danube' and he is disposed to feel sad if he comes to believe that Australia have lost their last cricket match. If this version of RTD\* is true then someone like Steve is possible. But this is counter intuitive: if we came across Steve and he claimed to have a desire that he has some chocolate then we would be unlikely to believe him: we might think that he was making a mistake about what it is that he wants. Conversely, if we saw Steve exhibiting these effects then we not take ourselves to be justified in attributing him with a desire that he has some chocolate. But this version of RTD\* entails that someone like Steve is possible, so it is false.

So the kinds of effects that can be the characteristic effects of having a particular desire cannot be completely unrestricted. Any plausible version of RTD\* must impose some constraint on what kinds of effects can be the characteristic effects of having a particular desire.<sup>104</sup> Such a constraint could

---

<sup>104</sup> Of course, someone could bite this bullet and insist that someone like Steve is possible despite this being counter intuitive. They might simply take this as an interesting and deep fact about desires that science has revealed about desire. Alternatively they might claim that this kind of dissociation between the object of a desire and the effects that having the desire

be imposed in one of two ways: on one hand it could entail that the characteristic effects of having a particular desire are a subset of those that in humans are the characteristic effects of having that desire, and that are related to the object of the desire, but that having a disposition to act in ways that you believe will bring about what you desire is not part of this subset. On the other hand it could entail that the characteristic effects of having a particular desire are any of those that in humans are the characteristic effects of having that desire, and that are related to the object of the desire, where this includes having a disposition to act in ways that you believe will bring about what you desire. Call a version of RTD\* that imposes the first kind of constraint 'RTD\*a', and call a version of RTD\* that imposes the second kind of

---

has is metaphysically possible, but is nonetheless nomologically impossible. That is, that given the natural laws in the actual world the typical effects of having a particular desire are necessarily related in some way to the object of that desire. For example, they would claim that it is just a fact that, given the natural laws if someone has a desire for chocolate then the typical effects on them of having this desire will be related to chocolate in some way.

However, these responses are inadequate here. The first option, to bite the bullet and insist that Steve is possible, and that the connection between the effects of having a particular desire and the object of that desire is merely contingent, leaves it unexplained why in actual cases the effects of having a particular desire are almost always related to the content of the desire. But if we think that explanatory power is a theoretical virtue (and presumably someone who is giving this kind of empirical analysis does) then we should not accept this kind of account when there are other versions that do not leave such a prominent fact about desire unexplained.

On the other hand if someone says that while Steve is metaphysically possible he is nonetheless nomologically impossible then they owe an account of the natural features or laws that make this the case. Otherwise they similarly leave it unexplained. It is no explanation to merely insist that there is a natural law that entails that the typical effects of having a particular desire must be related to the object of the desire. More worryingly, the kind of natural laws that would be required here would include psychophysical laws, that concerned connections between mental phenomena, the content of someone's desire and physical phenomena, any physical effects that are characteristic of having a particular desire. It is controversial whether there are any psychophysical laws at all. And even if there are

constraint 'RTD\*b'. RTD\*b entails that MN is true: it entails that if someone has a desire that *p* then they are disposed to act in ways that they believe will bring about *p*. On the other hand, RTD\*a is vulnerable to the Natural Kinds objection.

According to RTD\*a it is necessary and sufficient for having a desire that *p* that someone has the capacity to perceptually or cognitively represent that *p* to initiate the production of reward signals that cause characteristic effects such as disposing them to have experiences of pleasure if they come to believe that *p* has come about, altering their perceptual sensitivity for discriminating that *p*, disposing them to make positive evaluations of *p*, and so on. But also according to RTD\*a it is not necessary for having a desire that *p* that someone has the capacity to perceptually or cognitively represent that *p* to initiate the production of reward signals that cause them to be disposed to act in ways that they believe will bring about *p*. Yet, if what Schroeder claims is the best current information from neuroscience is true then there is no unified state that we have that has all and only the effects that, according to RTD\*a are necessary and sufficient for having a desire. Recall, according to this neuroscientific evidence, in humans it is the reward centre, the VTA/SNpc that initiates the reinforcement learning processes that have effects on dispositions to have experiences of pleasure, sensory capacities, intellectual dispositions,

---

any it would be a matter for speculation what particular connections they can involve. So the prospect of giving the required explanatory account in terms of natural laws is poor.

and so on. But it is also the VTA/SNpc that initiates the reinforcement learning processes that have effects on dispositions to act. According to RTD\*a it is necessary and sufficient for having desires that someone has those parts of the VTA/SNpc that initiate the reinforcement learning processes that have the first kinds of effects, and not necessary that they have the parts that initiate the reinforcement learning processes that have the second kind of effect, on dispositions to act. So RTD\* says that when a human has a desire this consists in their being in a state of some particular parts of the VTA/SNpc. But this is not a unified state. The VTA/SNpc might be considered to be a distinct neurological structure by current neuroscience, and states of the VTA/SNpc might be considered to be unified states. But this does not mean that states of some of its parts are together a unified state: it is more plausible that the states of only some, but not all, of the parts of a unified state do not together comprise a unified state. Consider, for example a car at time *t*. Intuitively the car is in a unified state at *t*. But also intuitively the state at *t* of each of its left rear tire, its steering wheel, its passenger side head-rest and its exhaust pipe taken together do not comprise a unified state, in the sense of being unified that is relevant here.<sup>105</sup> <sup>106</sup> So according to RTD\*a there is no

---

<sup>105</sup> That is, according to Schroeder, something that is suitable for meaningful scientific investigation [recall, (Schroeder 2004, pp. 5-7)].

<sup>106</sup> Of course, there are difficult and controversial mereological issues here. However, I will not attempt to address these: the example and the intuitions seem to support the distinction between unified and non-unified states in the sense that is relevant to the discussion.



unified state that we are in when we have desires. Yet, the intuitive assumption underlying the Natural Kinds objection is that it is necessary that if someone has a desire then they are in a unified state: so if RTD\*a is true then we do not have desires. But this is just the conclusion of the Natural Kinds objection, so RTD\*a does not provide a way of responding to that argument. It cannot be the version of reward theory that follows from Schroeder's response to that objection: Schroeder must intend his theory to be understood as RTD\*b.

Now consider RTD\*b: according to this version of reward theory it is necessary and sufficient for having a desire that *p* that someone has the capacity to perceptually or cognitively represent that *p* to initiate the production of reward signals that cause characteristic effects such as disposing them to have experiences of pleasure if they come to believe that *p* has come about, altering their perceptual sensitivity for discriminating that *p*, disposing them to make positive evaluations of *p*, and disposing them to act in ways that they believe will bring about *p*. This version of reward theory is not vulnerable to the Natural Kinds objection: according to the neuroscientific evidence, in humans it is the VTA/SNpc that initiates such reinforcement learning processes, and states of the VTA/SNpc are unified states. So according to RTD\*b when we have a desire we are in a state of the VTA/SNpc (as opposed to a state of some but not all of its parts as with RTD\*a), so we are

in a unified state. Hence, RTD\*b is consistent with the assumption underlying the Natural Kinds objection: this version of reward theory allows us to avoid the conclusion that we do not have desires. Yet it claims that it is a necessary condition for having a desire that someone is disposed to act in ways they believe will bring about what they desire. So RTD\*b entails that MN is true. Of course, it makes a stronger claim than MN: it also claims that this disposition must be brought about in a particular way, by reinforcement learning processes. But it entails that someone with a desire has this disposition nonetheless. So if this is the version of reward theory of desire that follows from Schroeder's response to the Natural Kinds objection then MN follows from Schroeder's response as well.

It seems that there are relations between reward and dispositions to act just as there are relations between desire and reward and between desire and dispositions to act. Having something that is a reward for you, in Schroeder's sense of the term, seems to entail having a disposition to act in ways that you believe will bring it about. So the most plausible versions of Reward Theory of Desire are not alternatives to motivational theories of desire but are themselves versions of motivational theory: they entail that MN is true.

## **5.9 – Conclusion.**

According to what I have called ‘the Natural Kinds objection’, recently made by Timothy Schroeder, MN conflicts with the best current empirical evidence that we have from neuroscience. This objection appeals to a methodological assumption from empirical functionalism and to empirical evidence to a greater extent than the objections to MN that I have considered in earlier chapters. According to this assumption, a plausible analysis of desire must entail that whatever desires are they are the kind of thing that can be investigated by empirical science. Yet, Schroeder claims that the results of current best theories in neuroscience suggest that there are no suitable, empirical states that have all of the properties and relations that desires are commonly taken to have. So, if we are to avoid the conclusion that we never have any desires then we must accept that some of these properties and relations are not, in fact properties and relations of desire. In addition, Schroeder claims that this empirical evidence suggests that the relation between having a desire and having a disposition to act is among the relations that are not necessary for having a desire, so MN is false. He argues that the neural correlate of having something as a reward for you is a better candidate to be the empirical state that we are in when we have a desire than the neural correlate of having a disposition to act. Hence, according to Schroeder we

should analyse desire in terms of a relation with reward, his Reward Theory of Desire, RTD\*.

However, to the extent that the empirical evidence supports Schroeder's claim that a relation between desire and reward is necessary for having a desire it also supports MN, that a relation between desire and dispositions to act is necessary for having a desire. Moreover, Schroeder's analysis in terms of a relation with reward is implausible: it conflicts with our common sense understanding of desire to too great an extent. If the Natural Kinds objection entails that desires are the kind of state that Schroeder's analysis implies then it seems that we do not have desires after all. So, contrary to Schroeder's claims, empirical evidence suggests that the relation between having a desire and being disposed to act is necessary for having a desire: there is empirical support that MN is true.

## CONCLUSION

I will now briefly summarise what I have argued in the previous chapters and what I have attempted to show. I will then return to a number of issues that have arisen during the course of these arguments that I have bracketed or been unable to address at those points. Dealing with such issues adequately would go beyond the scope of this dissertation. Nonetheless they are not irrelevant to my interests here as they raise a number of important questions about desires. If our aim is to gain a proper understanding of desire then they need addressing at some point. Although I will not attempt to do so in this brief discussion I will highlight some that I take to be relevant, interesting, and important areas for future research.

My overall aim in the dissertation has been to defend a view about a relation between desire and dispositions to act I call ‘the Motivational Necessity of Desire’ (MN). MN is the following thesis:

(MN)            If  $S$  desires that  $p$  then  $S$  is disposed to act in ways  $S$  believes will bring about  $p$ .

According to MN, if Chris desires that he sees a film this evening and Chris believes that he can see a film by going to the local cinema at 8 pm then Chris is disposed to go to the local cinema at 8pm. If he also believes that he can see a film by going to the video rental shop then he is disposed to go to the video rental shop. MN is *prima facie* plausible, and it is a claim that has been both traditionally made about desires and that has been very influential in discussions about desire in philosophy. There are good philosophical reasons to accept it.

I discussed a number of these reasons in Chapter One, Motivations for Motivational Necessity. I highlighted the prevalence of the view through a range of fields of philosophy and how it is presupposed by certain important debates in those fields. I then discussed some of the theoretical benefits of MN for work in philosophy of mind in particular. MN provides a plausible way of answering at least two traditional and persistent problems in philosophy of mind: that is, how to give an analysis of motivation and how to give an analysis of the relation between a person and the intentional objects of their mental states like desires that are consistent with naturalism. In addition, there are explanatory benefits for a theory of desire if it accepts MN. It provides the most straightforward explanation of a prominent fact about desire, that there is a regular correlation between having a desire for something and acting in ways that you believe will bring it about. Other features that

standard theories of desire attempt to analyse desire in terms of, like experiences of pleasure, evaluations, and rewards, must make further claims and arguments, in addition to their claim that desire is related to that particular feature, in order to explain this fact.

So there are good reasons to accept MN. Nonetheless, a number of seemingly persuasive objections have been made to the view. I addressed each of what I take to be the strongest objections that have been made to MN in turn in the Chapters Two to Five. In Chapter Two I defended MN against Galen Strawson's objection based on his counter example of the Weather Watchers. Strawson argued that it is possible that there are creatures like the Weather Watchers that can have desires but cannot have intentions. But he claimed that it is necessary for having a disposition to act that someone can have intentions. So if the Weather Watchers can have desires but cannot have intentions then they also cannot have dispositions to act: they are a counter example to MN. However, I argued that the case was under-described. I argued that none of the ways in which it could be coherently filled out that do not beg-the-question against MN, are cases in which the Weather Watchers both have desires and cannot have intentions. If the case is properly described in a way in which it is plausible that the Weather Watchers cannot have intentions then it is also plausible that they cannot have desires. On the other hand, if the case is properly described in a way in which it is plausible that the

Weather Watchers do have desires then it is plausible that they also have intentions to act and dispositions to act. Hence, the Weather Watchers are not a counter example to MN.

Next, in Chapter Three I discussed further objections that appealed to putative counter examples: these are examples of desires for particular things that someone cannot be disposed to act in ways that they believe will bring about. I considered three kinds of such desires: first, what I called ‘self-passive desires’, like a desire that your child passes their exams on their own. These are cases in which someone desires that they have something that they do not bring about by their own actions. Second, what I called ‘necessity desires’, like a desire that there are infinitely many prime numbers. These are cases in which someone desires something that they believe is necessarily the case. Finally, what I called ‘impossibility desires’, like a desire that I have a square circle. These are cases in which someone desires something that they believe necessarily cannot be the case. The objection in each case is that it is not plausible that someone who has such a desire is disposed to act in ways that they believe will bring about what they desire. The idea is that at least those people who are rational cannot believe that there are any ways that they can act that will bring about what they desire: hence they cannot be disposed to act in any such ways. But it is plausible that someone who is rational can have a desire of one of these kinds. So they are a counter example to MN:



they are cases in which someone both has a desire and is not disposed to act in ways that they believe will bring about what they desire.

However, I argued that none of these cases is a successful counter example to MN. In the cases of self-passive desires this is because of a misunderstanding about what the object of these desires actually is, and so a misunderstanding about the disposition that is entailed by MN. The disposition to act that people with a self-passive desire cannot have, according to the objection, is not a disposition that MN claims they will have if they have a self-passive desire. So these cases are not counter examples to MN. In addition, it is possible for someone who is rational to believe that there are ways in which they can act to bring about what the actual object of a self-passive desire is, properly understood, and it is plausible that in these examples the person with a self-passive desire is in fact disposed to act in these ways.

I argued that the cases of necessity desires and impossibility desires similarly fail to provide counter examples to MN: again this is because of a misunderstanding of the object of these desires. They are particular cases of a general problem about how we can have mental states that represent one and not another of two propositions that are necessarily equivalent. As the propositional objects of necessity desires are necessarily true, and the propositional objects of impossibility desires are necessarily false then all

necessity desires appear to have necessarily equivalent propositional objects, and all impossibility desires appear to have necessarily equivalent propositional objects. Yet all necessity desires do not appear to be the same, and all impossibility desires do not appear to be the same. Because of this general ‘problem of equivalence’, as Stalnaker calls it, it is not clear what the propositional objects of necessity desires and impossibility desires actually are. And if this is not clear then we cannot know what disposition to act is entailed by having such desires according to MN, and cannot know if someone with such a desire does not have that disposition. So these cases do not show that it is possible to have a desire and not be disposed to act in ways you believe will bring about what you desire. Neither self-passive desires, necessity desires nor impossibility desires are counter examples to MN.

My argument in Chapter Four was defensive again but if successful then I take it to provide positive support for MN. It does this in two ways: first, it provides independent motivation for thinking that MN is true, and second, it justifies the claim I made in Chapter One concerning the explanatory advantages of MN for an analysis of desire. I defended a particular way of analysing the difference between belief and desire. I argued that the most plausible way of analysing this is in terms of the different directions-of-fit of the respective kinds of state, where having a particular direction-of-fit consists in having a particular functional role, a particular set of

properties and relations. A plausible version of this kind of functional role DOF account of the difference between beliefs and desires is given by Michael Smith: Smith analyses the particular directions-of-fit of belief and desire in terms of their respective relations with conflicting perceptions and with dispositions to act, among others. I defended Smith's DOF account against a pair of objections, made by Sobel & Copp and by Humberstone. They argued, in different ways, that Smith's DOF account is uninformative: they claim that the difference between the properties it attributes to beliefs and desires is not a difference that is constitutive of the distinction between these kinds of states. So they claim that Smith's DOF account cannot be the right analysis of direction-of-fit. However, I argued that neither objection is successful: Sobel & Copp appeal to an implausible and overly-demanding claim about what is necessary for an analysis to be informative. Humberstone does not take account of the detail that is provided by Smith's DOF account in the characterisations it gives of the different directions-of-fit of belief and desire. These characterisations provide exactly the kind of information that Humberstone claims is necessary for this kind of analysis to be informative. So neither objection is successful.

If something like Smith's DOF account is correct then, as I said above, this supports MN in two ways: on one hand it is independent support for MN. It is the best way of analysing the difference between beliefs and desires, and it

is intuitive that there is such a difference. But according to the account having a desire that  $p$  entails having a disposition to act in ways that you believe will bring about  $p$ . So this intuitive difference between beliefs and desires appears to entail MN. On the other hand, Smith's DOF account justifies what I earlier claimed was an explanatory benefit of MN: that it provided the best explanation of the fact that having a desire is regularly correlated with acting in ways that you believe will bring about what you desire. This might be challenged if someone thought that the correlation that we observe, and that needs explaining is not between having a desire and acting but rather between having a complex of beliefs and desires, and acting. But if Smith's DOF account is true then having a belief does not entail having a disposition to act. So even if the correlation that needs explaining is between having a complex of beliefs and desires, and acting then, if Smith's DOF account is true this correlation is also best explained in terms of the properties and relations of the desires in such complexes of mental states. In other words, Smith's DOF account suggests that MN has explanatory benefits for a theory of desire with respect to explaining this prominent fact about desire, the regular correlation between having a desire and acting, as I had earlier claimed.

Finally, in Chapter Five I addressed what I called the 'Natural Kinds objection', made by Timothy Schroeder, that MN is incompatible with the best empirical evidence. According to Schroeder it is both assumed by our

everyday understanding of desire, and is a plausible constraint on an informative analysis of desire that, whatever a particular analysis claims that desires are, they must be the kind of thing that are appropriate for scientific investigation: in this particular case, having a desire must consist in being in a unified state. But, he claims that the best current neuroscientific evidence suggests that there are no unified states that we have that have all of the properties and relations that are commonly attributed to desire in our everyday understanding of it. In addition, Schroeder claims that the most plausible candidate for being a desire from among the unified states that we have is one that does not have a necessary connection with dispositions to act: rather it is a kind of state that is connected with something being a reward for you. So, according to Schroeder the best empirical evidence suggests that desire is related with reward and is not related with dispositions to act, so MN is false.

However, I argued that, contrary to Schroeder's claim, this empirical evidence supports MN to the same extent that it supports his own Reward Theory of Desire, that analyses desire in terms of a relation with reward. In addition, I argued that this kind of reward theory has implications that conflict with a number of intuitions that we have about desire. So it is not a plausible analysis of desire. In light of this it is reasonable to take the empirical evidence to suggest that desire is related with dispositions to act, rather than with reward if they are supported by it to an equal extent. So MN is not

incompatible with the best empirical evidence, as Schroeder claims: rather it appears to be supported by this evidence.

If my arguments are correct then none of these objections to MN succeed. And if these are the strongest objections to MN, as I take them to be, then in light of the *prima facie* plausibility and the theoretical benefits of this thesis we should accept it.

My aims in this dissertation have been fairly modest, and the arguments I have given mainly defensive. I have restricted myself to addressing objections to MN alone, the view that there is a necessary connection between having a desire and being disposed to act in particular ways. And I have merely attempted to show that these objections are unsuccessful against that particular view. I have not, in addition to this, attempted to give a theory of desire, a set of necessary and sufficient conditions for having a desire. Neither have I attempted to defend some related theses, for example that make claims about what kind of relation this connection between desire and dispositions to act is, or why they are connected in this way. But there are a number of claims that are often made about such things that are both interesting in themselves and are related specifically to the connection between desire and dispositions to act that MN is concerned with. In the rest of the conclusion I will briefly discuss some of these.

First, as I have noted at a number of points throughout the thesis, desires are often taken to have certain normative properties and relations. For instance, some people claim that desires can provide reasons to act in particular ways [see for example Stampe (1987)], while others claim there are constraints on what can be the proper objects of desire, that is what it is rational to desire [see for example Anscombe (1957)]. Both of these claims concern what someone has reason to do, or what it is rational to do: they say that, insofar as someone is rational then they should act in a certain way, or they should not desire a particular kind of thing. So these are claims about normative properties and relations of desire.

Consider the first of these, the claim that having a desire can provide someone with a reason to act in a particular way. This is what Stampe calls the '*per se* authority of desire' (Stampe 1987, p343). According to Stampe, if Adam desires to go to Peterborough, and he believes that he can get to Peterborough by getting on the 11.10 from Euston station then he has a pro tanto reason to get on the 11.10 from Euston station. Moreover, Adam has this reason merely in virtue of having this particular desire. Stampe gives the following example to illustrate the claim:

Consider two people who do exactly the same thing, between whom there is no difference save that one of them did the thing because that

one wanted to do it. (Suppose they both spent their weekend studying German, with the intention of learning the language.) Beyond that neither of them has any belief about the act that would explain their curious behaviour; neither believes that there is any point in his learning German. Then the one who has no desire to do this thing has absolutely no reason to do it (none in view of which he does it), and the act is utterly irrational. But the act of the one who wants to do it is not *utterly* irrational. He does at least have a reason – something of a reason, surely – to do it, for it is something that he wants to do. And if he has a reason then there is a reason for him to do it. What can this be but the only thing that he has that the other party lack? That is: the desire itself. The *desire*, it may be, is irrational, but it does not follow that the act that it inspires is irrational. For the action is done for a reason: the desire the agent has to do it. Desire, it seems, has an extraordinary authority: if there is no reason to do a thing, my *intending* to do it is no reason to do it, and neither is my *believing* that it would be a good idea to do it, if in fact it would not. But even then, the fact that I *want* to do it *is* a reason to do it (Stampe 1987, p344).

Note the particular claim that is being made here. Stampe is not saying merely that there are normative facts about desire. Rather he is saying that there is a



particular normative fact about desire that is not true about other kinds of mental state, like beliefs, intentions and so on. That is, that if someone has a desire for something they thereby have a reason to act to try to get what they desire.

Note also that it is a restricted claim: it is merely a *pro tanto* reason that someone has in virtue of having a desire, a reason that can be over-ridden by other reasons they might have (say, on the basis of other desires, or on the basis of other facts about themselves or the situation they are in). Having a desire does not thereby provide someone with an all-things-considered reason to act to try to get what they desire.

It is plausible, and widely agreed that having a desire can provide someone with such a reason to act.<sup>107</sup> And this normative fact about desire seems particularly relevant to what I have been arguing here. It is closely related to the thesis that I called MN(rat) in the Introduction when distinguishing it from MN. Recall, it is the following, restricted thesis:

[MN(rat)]      If *S* desires that *p* then *S* is disposed to act in ways *S* believes will bring about *p*, insofar as *S* is rational.

---

<sup>107</sup> Although, see for example Dancy (1999), Scanlon (1998), etc for disagreement.

Whereas MN is a claim about all things that have a desire, MN(rat) is a claim about all of the things that are rational that have a desire. To put this another way, MN(rat) says that if someone has a desire then they have a reason to be disposed to act in ways that they believe will bring about what they desire. Someone can have a reason to do something only to the extent that they are rational: to say that someone is irrational is to say that they are not sensitive to reasons. And it would be very peculiar if someone could have a reason to be disposed to act in a particular way but not have a reason to act in that way.<sup>108</sup> Although I have concentrated on MN rather than MN(rat) in the dissertation, as I noted at a number of points MN(rat) is weaker than, and is entailed by MN. So if this normative fact is relevant to MN(rat) then it is also relevant to MN.

Now consider the second kind of normative claim about desire that I mentioned above: the claims that there are certain constraints on the proper objects of desire. This idea might be expressed by saying that there are

---

<sup>108</sup> Although, there are likely to be exceptions to this: situations in which someone has a reason to be disposed to act in a certain way but does not have a reason to act in that way. Perhaps Kavka's 'toxin problem' can be thought of like this. This is a case in which you have a reason to sincerely intend to drink some a non-lethal poison in the future despite the fact that (you know) at that time you will have no reason to actually drink the poison. If having a disposition to act entailed having an intention to act (see below, and Chapter Two, section 2.3 for more on this) then this might be a case in which you have a reason to have a disposition to act in a particular way but do not have a reason to act in that way. Nonetheless, the possibility of such cases would not show that Stampe's claim and MN(rat) are unrelated. Stampe's claim is that having a desire can thereby give someone a reason to act, not that it entails having such a reason. So if there are situations in which it would be rational to act in the way you are disposed to act then this normative fact will follow from MN(rat).

correctness conditions for the objects of desire, such that someone is criticisable if they have a desire for something that violates these correctness conditions. Perhaps violating such conditions can be understood in terms of the object of someone's desire lacking certain properties that, according to claims about these constraints a desire should have. According to Anscombe for example, if something is an appropriate object of desire then it has some desirable characteristic (either intrinsically or in virtue of being a means to obtaining something that has some desirable characteristic intrinsically). That is, insofar as someone is rational then they cannot have a desire for something that they believe does not have any desirable characteristics. So she says that insofar as someone is rational then they cannot have a desire for a saucer of mud, or a desire for a twig of mountain ash (where these things are not being desired as means to obtaining something else). Anscombe takes it to be inconceivable that this kind of object has any desirable characteristics intrinsically. So if someone is rational, and so can only have desires for things that do not violate the correctness conditions on the objects of desire then they cannot have a desire for a saucer of mud. It is plausible that there are constraints like this on the proper objects of desire. For example, it is intuitive that there is something wrong with having a desire to inflict pain on someone merely for the sake of inflicting pain [rather than, for example, to punish them for some wrongdoing or for the pleasure someone might get from inflicting

pain, (which might be wrong because of other, for example ethical, constraints rather than because of constraints on the objects of desire)]. But it is controversial what exactly the constraints actually: whether they include this constraint described by Anscombe, that the proper objects of desire have some desirable characteristic, and if so whether there are others in addition to it, what these are, and so on.

Although this issue about the constraints on the proper objects of desire has not come up itself in the arguments in the previous chapters, it is related some of issues that have been relevant. For example, in Chapter Three, section 3.10 I considered a response that is sometimes given to the putative counter examples to MN of necessity desires and impossibility desires, desires like a desire that there are infinitely many prime numbers and like a desire that I have a square circle respectively. Recall, that response denies that it is possible to have these kinds of desire: it denies that someone can genuinely desire something that they believe is necessarily the case or that they believe necessarily cannot be the case. Rather, according to this response, in cases where it seems as though someone has a desire for such an object they in fact have a different kind of pro-attitude towards it, like a hope or a wish. That is, the putative examples in which someone is said to have a desire that there are infinitely many prime numbers are really cases in which they hope that there are infinitely many prime numbers, for example. So they are not counter

examples to MN as this is a thesis about desire and not about other kinds of pro-attitude as well. The motivation behind this response is plausibly that things that are believed to be necessarily the case or are believed to be impossible are not proper objects of desire: they violate some constraint on the proper objects of desire. So insofar as someone is rational then they cannot have a necessity desire or an impossibility desire. How plausible this response is therefore depends on what the particular constraints on the proper objects of desire actually are.

The fact that desires have such normative properties and relations is often thought to be a problem for MN.<sup>109</sup> The idea is that MN is a claim about a natural relation of desire, and that normative properties and relations cannot be derived from natural ones. So MN is thought to be inconsistent with desire having normative properties and relations, which would entail that MN is false. But this is a mistake, for a number of reasons. First, MN is merely a claim about a necessary condition for having a desire. It is not a claim about necessary and sufficient conditions. So even if these normative facts about desire cannot be explained in terms of the relation that MN describes, this does

---

<sup>109</sup> Recall, for example the objection to Smith's DOF account raised by Zangwill (1998) among others, that I mentioned in Chapter Four, section 4.10, but did not discuss in detail. According to Zangwill, Smith's DOF account is uninformative because it attempts to analyse the difference between beliefs and desires in terms of the natural properties and relations of each kind of state. He claims that this difference is a difference in the normative properties and relations that each kind of state has, and that normative facts cannot be explained in terms of natural facts. Hence, the differences described by Smith's DOF account are not the ones that distinguish between belief in particular and desire in particular.

not entail that they cannot be explained in terms of whatever other conditions are necessary for having a desire. So MN is not inconsistent with desires having normative properties and relations even if it does not entail that desire has them.

Moreover, it has not been established that MN cannot explain these normative facts about desire. People who raise the normative facts about desire as a problem for views like MN often appeal to things like Moore's 'Open Question' argument, and the slogan that 'you cannot derive and ought from an is'.<sup>110</sup> But although this slogan is often taken as a dogma, it is controversial whether we should accept Moore's argument, and controversial what it actually shows if it is sound.<sup>111</sup> So, we should not merely accept without further argument the claim that normative facts cannot be derived from natural facts: we should not accept this premise in these objections that appeal to the natural facts about desire.

Note, also, that these normative facts about desire are a problem for all standard kinds of theory of desire, irrespective of the properties or relations that they claim are necessary conditions for having a desire, just as they are for MN. This is obvious in the case of hedonic theories and reward theories of desire. These kinds of theories respectively attempt to analyse desire in terms of relations between desires and dispositions to have experiences of pleasure

---

<sup>110</sup> Kripke (1982) also discusses this in the context of following a rule.

and relations between desires and something being a reward, where this is usually understood in terms of the conditioning, or reinforcement learning effects of the object of desire [see, for example, Schroeder (2004)]. But these are natural properties and relations, just like dispositions to act described by MN. So if MN cannot account for the normative facts about desire because it is a claim about a natural relation of desire then hedonic theories of desire and reward theories of desire cannot account for these facts either. But, perhaps surprisingly, these normative facts are also a problem for evaluative theories of desire, that attempt to analyse desire in terms of making positive evaluations of the object of desire. It might seem that evaluative theories have an advantage over the other kinds of theory with respect to this problem. Because they attempt to analyse desire in terms of evaluations they presuppose that desire has relations with facts about value, which are normative facts. So they do not face the problem of explaining how desire can have any normative properties or relations at all. However, recall what it is that needs accounting for. It is not merely that desires have normative properties and relations, that there are some normative facts about desire. What stands in need of explanation is that desire has the particular normative properties and relations that it does, as opposed to having the same ones as belief, for example. What Stampe was concerned to show with his example of the two people studying German was

---

<sup>111</sup> See for example, Joyce (2006) for discussion, especially the Introduction.

that it is only having a desire, and not having any other kind of mental state, that can thereby provide someone with a reason to act. Similarly, it is not especially interesting that desires have correctness conditions: it is uncontroversial that mental states can have correctness conditions, no one takes this to be particularly problematic in the case of belief, for example. What is most interesting and controversial is that desires have the particular correctness conditions that they do, for example that insofar as someone is rational someone cannot desire something that they take to have no desirable characteristic.<sup>112</sup> Indeed, it seems that even those who appeal to the normative facts about desire when making an objection to MN are appealing to the fact that desires have the particular normative properties and relations that they do. Consider Zangwill's objection to Smith's DOF account once again. He claims that this account is inadequate because it fails to give an account of what is constitutive of the difference between beliefs in particular and desires in particular. That is, their difference in normative properties and relations. So this objection depends on their being a difference between the normative facts about each kind of state. Zangwill's objection against Smith's DOF account for example, could not even get started if desires had the same normative properties and relations as beliefs. So what stands in need of explanation is

---

<sup>112</sup> This is not to claim that it is not controversial whether or not there are correctness conditions for desire. Instrumentalists about desire, for example, deny that there are any normative constraints on the possible objects of desire [see for example Fehige (2001) for discussion].



that desires have the particular normative properties and relations that they do. But evaluative theories of desire do not have a clear advantage over other kinds of theory with respect to explaining this. The fact that they analyse desire in terms of positive evaluations of the objects of desire might thereby entail that there are normative facts about desire. But it does not thereby entail that there are the particular normative facts about desire that, in fact, there are, and that are what stands in need of explanation. So, in order to explain this evaluative theories must make further claims about the feature that they analyse desire in terms of, just as the other kinds of theory of desire that analyse desire in terms of natural properties and relations must do. So if there are normative facts about desire and this is a problem for MN then it is also a problem for all of the standard kinds of theory of desire. However, this is not to deny that these normative facts about desire cannot be explained in a straightforward way by MN, in contrast to the way in which the regular correlation between having a desire and acting in a way you believe will bring about what you desire can be straightforwardly explained. If there are normative properties and relations of desire, as seems plausible, then there are interesting questions about what exactly they are: and there are interesting questions about whether, and how they might be explained in terms of MN. Both kinds of question deserve further attention.

A different issue that I left unanswered earlier concerns the relations between desires and intentions and between intentions and dispositions to act. Recall, from Chapter Two, Strawson claimed that it is a necessary condition for having a disposition to act that you can have an intention to act. In addition he claimed that it is possible that someone can have desires but cannot have intentions, attempting to show this with the Weather Watchers example. If both of these claims were both true then MN would be false, it would be possible for someone to have a desire yet not have a disposition to act, because they cannot have an intention to act. In Chapter Two I responded to this objection by arguing that the Weather Watchers do not show that it is possible that someone can have desires but cannot have intentions, and that there is reason to think that it is not possible. I granted Strawson his assumption that having an intention is necessary for having a disposition to act. Indeed it is very plausible that there is some kind of relation between intentions to act and dispositions to act, as I discussed in Chapter Two, section 2.3. But does Strawson describe it correctly? Recall from that section, I discussed a reason someone might have to deny Strawson's claim. If someone thought that cases like those described by Hursthouse (1991), in which someone performs an action in the grip of, or as the expression of, an emotion are cases in which they do not have an intention then it seems that Strawson's claim is false. Nonetheless, it seems that even someone like Hursthouse would accept that

cases in which someone does have an intention to perform a particular action and that this causes them (in the right way) to make the appropriate bodily movements are cases in which they have acted. It is plausible that even someone like Hursthouse who might deny that having intentions is necessary for having dispositions to act might accept that having an intention is sufficient for having a disposition to act. Might Strawson have mis-described the direction of the entailment relation between intentions to act and dispositions to act?

Note the implications this would have for Strawson's objection. If my argument in Chapter Two is correct then if someone can have a desire then they can have an intention to act. In addition, if having an intention to act is sufficient for having a disposition to act, then MN follows by logical implication: MN is the conclusion of the following hypothetical syllogism:

- 1) If  $S$  desires that  $p$  then  $S$  has an intention to act in ways that  $S$  believes will bring about  $p$ .
- 2) If  $S$  has an intention to act in ways  $S$  believes will bring about  $p$  then  $S$  is disposed to act in ways  $S$  believes will bring about  $p$ .
- C) If  $S$  desires that  $p$  then  $S$  is disposed to act in ways  $S$  believes will bring about  $p$ .

In fact, the problem for Strawson is even more acute. Even if having an intention is not, after all, related to each of desires and dispositions to act in these ways, Strawson must deny that there is anything that has these relations: Strawson must deny that there is any property or relation that is both a necessary condition for having a desire and a sufficient condition for having a disposition to act such that it can be an intermediary between desires and dispositions to act in an analogous way. Of course, Strawson will deny this: as I described in Chapter One, section 1.5, Strawson seems to deny that there are any properties or relations that are necessarily connected with desire. He seems to accept a primitivist theory of desire. But note a further implication of this. If someone denies that there is any property or relation that is an intermediary between desires and dispositions to act in this way then they are effectively denying that there is any interesting explanation, in terms of the properties of desire, of the fact that people often act in ways that they believe will bring about their desire. They must claim that it is a merely contingent fact that there is a regular correlation between having a desire and acting to try to get what you desire. This is counter intuitive: it is reason to think that there is some such property or relation that is both entailed by having a desire and entails having a disposition to act. In which case, MN is true.

There are many other issues about desire in addition to these that I have not even mentioned either in the body of the dissertation or in this

conclusion. But at least these issues raised here, that are closely related to my defense of MN are things that are in need of, and deserve further investigation.

## BIBLIOGRAPHY

- Adams, E. W. 1987: 'On the Meaning of the Conditional', *Philosophical Topics* (Spring 1987), pp. 5-22.
- Anscombe, G.E.M. 1957: Intention (Basil Blackwell: Oxford)
- Appiah, A. 1987: "'If' Again", *Analysis* O87, 47, pp 193-199.
- Aristotle 1961: De Anima (trans. Ross, D.) (Clarendon Press: Oxford).
- Armstrong, D. M. 1968: 'The Headless Woman Illusion and the Defence of Materialism', *Analysis*, D68, 20, pp. 48-49
- 1961: Perception and the Physical World (Routledge & K. Paul: London)
- Audi, R. 1994: 'Dispositional beliefs and dispositions to believe' *Nous*, 28, pp. 419-434
- 1993: Action, Intention, and Reason (Cornell Univ. Press: Ithaca)
- 1991: 'Intention, Cognitive Commitment, and Planning', *Synthese*, 91, pp. 361-378

- 1977: 'The Rational Assessment of Emotions',  
*Southwestern Journal of Philosophy*, Sum 77, 8, pp115-119.
- 1973: 'Intending', *Journal of Philosophy*, Jl 73, 70, pp. 387-403
- Austin, J. L. 1964: Sense and Sensibilia (OUP: New York)
- 1961: Perception and the Physical World (Routledge and Kegan Paul: London)
- Ayer, A.J. 1940: The Foundations of Empirical Knowledge (MacMillan: New York)
- 1936: Language, Truth, and Logic (Gollancz: London)
- Ayers, M. 2002: 'Is Perceptual Content Ever Conceptual?',  
*Philosophical Books*, Ja 02, 43(1), pp. 5-17
- Badcock, C. 2004: 'Emotion Versus Reason as a Genetic Conflict', in  
Emotion, Evolution, and Rationality [Evans, D. and Cruse, P. (ed)] (OUP: New York)
- Baker, L. R. 1993: 'What Beliefs Are Not', in *Naturalism: A Critical Appraisal* [Wagner, S. & Warner, R. (eds)] (Univ. of Notre Dame Press)
- 1987: Saving Belief (Princeton University Press: Princeton)

- Bayne, T. 2005: 'Putting the experience of acting in its place' draft  
[http://www.phil.mq.edu.au/staff/tbayne/direction\\_of\\_fit.pdf](http://www.phil.mq.edu.au/staff/tbayne/direction_of_fit.pdf)  
 (accessed 14/7/05)
- Bayne, T. and Pacherie, E. 2005: 'In Defense of the Doxastic Conception of Delusions',  
*Mind and Language*, Ap 05, 20(2), pp. 163-188.
- Bealer, G. 1997: 'Self-Consciousness' *The Philosophical Review* Vol  
 106, No.1 (Jan, 1997)  
 1982: Quality and Concept (Clarendon Press: Oxford)
- Ben-Ze'ev, A. 2003: 'The Logic of Emotions', *Philosophy* 52(supp.), pp.  
 147-162.  
 1997: 'Appraisal Theories of Emotions', *Journal Of  
 Philosophical Research*, Ap 97, 22, pp. 129-143
- Bermudez, J-L 2003: 'Non-conceptual content'; Stanford Encyclopedia of  
 Philosophy. Accessed Online 31/12/2005  
<http://plato.stanford.edu/entries/content-nonconceptual/>  
 2000: 'Naturalised Sense Data', *Philosophy and  
 Phenomenological Research*, 61, pp. 353-374
- Blackburn, S. 1998: Ruling Passions (OUP: Oxford).  
 1984: Spreading the Word (OUP: Oxford).



- Block, N. 1978. 'Troubles with Functionalism.' *Minnesota Studies in the Philosophy of Science* 9:261-325. Reprinted in *Readings in the Philosophy of Psychology* (MIT Press, 1980).
- Boghossian, P. 1989: 'The rule-following considerations.' *Mind* New series Vol. 98, No. 392 (Oct. 1989) pp. 507-549.
- Braddon-Mitchell 1996: Philosophy of Mind and Cognition  
D, and Jackson, F. (Blackwell: Cambridge, Mass).
- Bratman, M. 1999: Faces of Intention: Selected Essays on Intention and Agency (Cambridge University Press: Cambridge).
- 1993: 'Shared Intention', *Ethics*, O 93, 104(1), pp. 97-113
- 1991: 'Cognitivism and Practical Reason', *Ethics*, O 91, 102(1), p117-128
- 1987: Intentions, Plans and Reason (Harvard University Press: Cambridge, Mass.)
- 1984: 'The Two Faces of Intention'; *The Philosophical review*, Vol 93, No.3 (July 1984) pp. 375-405.
- Brentano, F. 1995: Psychology from an Empirical Standpoint [McAlister, L. (ed)] (Routledge: London)
- Brewer, B. 2004: 'Realism and the Nature of Perceptual Experience', *Nous Supplement*, 14, pp. 61-77
- 1999: Perception and Reason (OUP: New York)

- Brink, D. 1989: Moral Realism and the Foundations of Ethics.  
(Cambridge University Press: Cambridge).
- Broome, J. 2004: 'Reasons', in Reason and Value: Themes from the Moral Philosophy of Joseph Raz [Wallace, R.J. (ed)]  
(Clarendon Press: Oxford)
- 2001: 'Are Intentions Reasons? And how should we cope with Incommensurable Values?', Practical Rationality and Preference: Essays for David Gaultier [Morris, C. and Ripstein, A (eds.)] (Cambridge University Press: Cambridge)
- 1999: 'Normative Requirements'; *Ratio*, 12 (1999) pp.398-419.
- Byrne, A. 2003: 'Consciousness and Nonconceptual Content', *Philosophical Studies*, 113, pp. 261-274
- Carnap, R. 1937: 'Testability and meaning' in *Philosophy of Science*, III (1936) and IV (1937).
- Chalmers, D. 2006: 'Two-Dimensional Semantics.' In [(E. Lepore & B. Smith, (eds)] Oxford Handbook of the Philosophy of Language. (OUP: New York).

- 2006: 'Perception and the Fall from Eden' In [T. Gendler & J. Hawthorne, (eds)] Perceptual Experience. (OUP: New York)
- 2000: 'On Sense and Intension.' *Philosophical Perspectives* 16.
- 1996: The Conscious Mind: In Search of a Fundamental Theory (OUP: New York)
- Chisholm, R. 1989: 'Why Singular Propositions?' in Themes from Kaplan (OUP: New York).
- 1989: 'The Objects of Sensation: A Brentano Study', *Topoi*, Mr89, 8, pp. 3-8.
- 1984: 'The Primacy of the Intentionality', *Synthese*, O84, 61, pp. 89-110.
- 1982: The Foundations of Knowing (Univ. Minnesota Press: Minneapolis).
- Churchland, P. 1981. Eliminative materialism and the propositional attitudes. *Journal of Philosophy* 78:67-90. Reprinted in A Neurocomputational Perspective (MIT Press, 1989).
- Clark, A. 2002: 'Visual Experiences and Action: Are the Bonds too Tight?' *Philosophical Review*, 110, pp. 492-519

- Clark, R. 1981: 'Sensing, Perceiving, Thinking', *Grazer Philosophische Studien* 12, pp. 273-295
- Clarke, S.G. 1986: 'Emotional Rationality Without Cognitivism', *Dialogue*, Wint 86, 25, pp. 663-674
- Cohen, L. J. 1996: 'Do Beliefs Supervene on Degrees of Confidence?', in Connectionism, Belief, Cognition, and the Will [Meijers, A. (ed)] (Tilburg University Press)
- Cohen, S. 2005: 'Knowledge, Speaker and Subject', *Philosophical Quarterly*, Ap 05, 55(219), pp. 199-212.
- Copp, D. 1997: 'Belief, Reason, and Motivation: Michael Smith's, The Moral Problem' *Ethics* 108, no. 1 (1997): pp. 33-54.
- Coltheart, M. 2005: 'Conscious Experience and Delusional Belief', *Philosophy, Psychiatry, and Psychology*, Je 2005, 12(2), 153-157
- Crane, T. 2005: 'What is the Problem of Perception', *Synthesis Philosophica*, 20:2(40), pp. 237-264
- 2002: The Contents of Experience: Essays on Perception (Cambridge Univ. Press: Cambridge)
- 1992: 'The Nonconceptual Content of Experience', in The Contents of Experience [Crane, T. (ed)] (Cambridge University Press: Cambridge)

- 1988: 'The Waterfall Illusion', in Essays on Nonconceptual Content [York, H. (ed)], pp. 231-235, (Bradford: Cambridge, Mass.)
- Crick, P. 1976: 'Three Paradoxes of the Photograph', *British Journal of Aesthetics*, Sum 76, 16, pp. 268-271
- Currie, G. 2004: 'Imagination and Make-Believe'; The Routledge Companion to Aesthetics [Gaut, B. & McIver Lopes, D. (eds)] Second Edition (Routledge: London)
- Currie, G. & Ravenscroft, I. 2002: Recreative Minds (OUP: New York)
- Dancy, J. P. 2002: Practical Reality (Clarendon Press: Oxford)
- 1999: 'Motivation, Dispositions and Aims', *Theoria* vol. 65.2, pp. 144-55.
- 1993: Moral Reasons. (Blackwell: Oxford).
- Darwall, S. 1983: Impartial Reason (Cornell University Press: Ithica, NY).
- Davidson, D. 1980: Essays on Actions and Events (OUP: New York)
- Davies, M. 2001: 'Monothematic Delusions: Towards a Two-Factor Account', *Philosophy, Psychiatry, and Psychology*, Je 2001, 8(2-3), pp. 133-158.

- 1992: 'Perceptual Content and Local Supervenience',  
*Proceedings of the Aristotelian Society*, 92, pp. 21-45
- Dennett, D. 1983: 'Beyond Belief', in Thought and Object [Woodfield, A. (ed)] (OUP: New York), reprinted in The Intentional Stance (MIT Press: Cambridge, Mass. 1987)
- 1981: Brainstorms (MIT Press: Cambridge, Mass.)
- De Sousa, R. 2003: 'Paradoxical Emotion: On Sui Generis Emotional Irrationality', in Weakness of Will and Practical Irrationality [Stroud, S (ed)], pp. 274-297.
- 1987: The Rationality of Emotions (MIT Press: Cambridge, Mass.)
- Dretske, F. 1999: 'Machines, Plants and Animals: The Origins of Agency', *Erkenntnis*, 51(1), pp. 19-31
- 1988: Explaining Behavior: Reasons in a World of Causes (MIT Press: Cambridge, Mass.)
- Evans, G. 1982: The Varieties of Reference (Clarendon Press: Oxford)
- Fara, M. 2005: 'Dispositions and Habituals'; *Noûs* 39, pp.43-82.
- Fehige, C. 2001: 'Instrumentalism'; Varieties of Practical Reasoning [Millgram, E. (ed)] (MIT Press: Cambridge, Mass.)
- Fernandez, J. (Forthcoming): 'Self Knowledge, Rationality and Moore's Paradox'; *Philosophy and Phenomenological Research*.

Accessed online 31/12/2005  
<http://www.phil.mq.edu.au/staff/jfernandez/PPR@submission.pdf>

- Field, H. 1978: 'Mental Representation', *Erkenntnis* 13, pp. 9-61.
- Fine, C., Craigie, 2005: 'Damned if You Do; Damned if You Don't: the Impasse in Cognitive Accounts of the Capgras Delusion', *Philosophy, Psychiatry, and Psychology*, Je 2005, 12(2), pp. 143-151.
- Fisher, J. 1978: 'On Perceiving the Impossible', *British Journal of Aesthetics*, Wint 78, 18, pp. 19-30
- Foot, P. 1972: 'Morality as a System of Hypothetical Imperatives.' *Philosophical Review* (July 1972), 81(3):305-316.
- Forbes, G. 1989: 'Biosemantics and the Normative Properties of Thought'; *Philosophical Perspectives*, Vol. 3, Philosophy of Mind and Action Theory (1989) pp533-547.
- Frankena, W. 1958: 'Obligation and Motivation in Recent Moral Philosophy' in *Essays in Moral Philosophy*, [Melden, A.I. (ed)] pp. 40-81.
- Frege, G. 1879: 'Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens' (translated as 'Concept Script, a formal language of pure thought modeled upon that

- of arithmetic' by S. Bauer-Mengelberg) in From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931 [J. van Heijenoort (ed.)] (Harvard University Press: Cambridge, Mass. 1967)
- Garfield, J. 1988: Belief in Psychology: A Study in the Ontology of Mind (MIT Press: Cambridge, Mass.)
- Gibbard, A. 1990: Wise Choices, Apt Feelings: A Theory of Normative Judgment (Harvard University Press: Cambridge, Mass.).
- Goldie, P. 2004: 'Emotion, Reason, and Virtue', in Emotion, Evaluation, and Rationality [Evans, D. and Cruse, P. (ed)] (OUP: Oxford)
- Gombrich, E. H. 1960: Art and Illusion: A Study in the Psychology of Pictorial Representation (Phaidon Press: London)
- Gombrich, E. H. 1972: Art, Perception, and Reality (Johns Hopkins Press: Baltimore)
- Hochberg, J. and Black, M.
- Graham, G 2002: 'Behaviourism' Stanford encyclopedia of philosophy (<http://plato.stanford.edu/entries/behaviorism/#2>) accessed 3/2005
- Greenspan, P. 2003: 'Emotions, Rationality and Mind/Body', Philosophy 52(supp.), pp. 113-125.



- 1988: Emotions and Reasons: An Inquiry into Emotional Justification (Routledge: New York).
- Griffiths, P.E. 2003: 'Emotions', in Blackwell Guide to Philosophy of Mind [Stich, S and Warfield, T. (ed)] (Blackwell: London)
- 1997: What Emotions Really Are: The Problem of Psychological Categories (University of Chicago Press: Chicago)
- Haack, S. 1996: 'Reply to Commentators', *Philosophy and Phenomenological Research*, S96, 56(3), pp. 641-656.
- Hajek, A. 2005: 'My Philosophical Position Says 'p' and I Don't Believe 'p''; Moorean Absurdity: Essays on Content, Context and Their Collision (Green, M. and Williams, J. (eds.)) (OUP: New York)
- Hajek, A. & Pettit, P. 2004: 'Desire beyond Belief'; *Australasian Journal of Philosophy*, Vol. 82, No. 1. pp. 77-92.
- Hajek, A. & Stoljar, D. 2001: 'Crimmins, Gonzales and Moore'; *Analysis* 61.3 (pp 208-13).
- Hampshire, S. 1965: Freedom of the Individual (Harper and Row: New York)
- Hare, R. M. 1952: The Language of Morals. (OUP: Oxford).

- Harman, G. 1983: 'Rational Action and the Extent of Intentions', *Social Theory and Practice*, Sum-Fall 83, 9, pp. 123-142
- Hellie, B. 2005: 'Noise and Perceptual Indiscriminability'; *Mind*, Vol. 114, pp. 481-508.
- Helm, B. 2002: 'Felt Evaluations: A Theory of Pain and Pleasure', *American Philosophical Quarterly*, Ja 02, 39(1), pp. 13-30  
1994: 'The Significance of Emotions', *American Philosophical Quarterly*, O94, 31(4), pp. 319-331.
- Hintikka, J. 1962: Knowledge and Belief (Cornell University Press: Ithica)
- Hohwy, J. and Rosenberg, R. 2005: 'Unusual Experiences, Reality Testing and Delusions of Alien Control', *Mind and Language*, Ap 05, 20(2), pp141-162.
- Hookway, C. 2003: 'Affective States and Epistemic Immediacy', *Metaphilosophy*, Ja 03, 34(1-2), pp. 78-96
- Hume, D. 1975: An Enquiry Concerning Human Understanding [Selby-Bigge (ed.)] (Clarendon Press: Oxford)  
1965: A Treatise of Human Nature [Selby-Bigge (ed.)] (Clarendon Press: Oxford)

- Humberstone, L. 1997: 'Two types of Circularity' <http://www.jstor.org/cgi-bin/jstor/printpage/00318205/di982546/98p1074e/0-150.pdf?backcontext=results&dowhat=Acrobat&0-150.pdf>
- 1992: 'Direction of Fit' *Mind* 101, No. 401, pp.59-83.
- Hurley, S. 1998: Consciousness in Action (Harvard University Press: Cambridge, Mass.)
- Hursthouse, R. 1991: 'Arational Actions', *Journal of Philosophy*, 88, No. 2 (Feb. 1991), pp. 57-68.
- Jackson, F. 1996: 'Mental Causation: The State of the Art', *Mind*, 105, 419, pp. 377-413.
- 1986: 'What Mary Didn't Know', *Journal of Philosophy*, 83, pp. 291-5
- 1982: 'Epiphenomenal Qualia'; *Philosophical Quarterly* 32: 127-136.
- Jackson, F. and 1988: 'Functionalism and Broad Content', *Mind*, XCVII, 387, pp. 381-400.
- Pettit, P.
- James, W. 1884: 'What is an Emotion?', *Mind*, 19, pp. 188-204.

- Johnson, S. 1759: 'The History of Rasselas, Prince of Abbyssinia' in Rasselas and Other Tales [Gwin J. Kolb (ed)] vol. XVI of the Yale Edition of the Works of Samuel Johnson [Yale Univ. Press: New Haven (1990)].
- Johnston, M. 2004: 'The Obscure Object of Hallucination', *Philosophical Studies*, J1-Ag-S 04, 120(1-3), pp. 113-183
- Jones, K. 2004: 'Emotional Rationality as Practical Rationality', Setting the Moral Compass: Essays by Women Philosophers [Calhoun, C. (ed)], pp. 333-352 (OUP: Oxford).
- Joyce, R. 2006: The Evolution of Morality (MIT Press: Cambridge, Mass.)
- Kaplan, D. 1978: 'Dthat', in Syntax and Semantics, Vol. 9 [Cole, P. (ed)] (Academic Press: New York). Reprinted in The Philosophy of Language, [Martinich, A. P. (ed)] (OUP: Oxford).
- Katz, L.D. 2004: 'Review of Three Faces of Desire'; *Notre Dame Philosophical Reviews* (accessed online (19/12/2005) <http://ndpr.nd.edu/review.cfm?id=3861>)
- Kavka, G. S. 1983: 'The Toxin Puzzle'; *Analysis*, Vol. 43, No. 1 (Jan. 83), pp. 33-36.

- Kelly, S. 2001: 'The Nonconceptual Content of Perceptual Experience: Situation Dependence and Fineness of Grain', *Philosophy and Phenomenological Research*, My 01, 62(3), pp. 601-608
- Kerner, G.C. 1970: 'Passions and the Cognitive Foundations of Ethics', *Philosophy and Phenomenological Research*, D 70, 31, pp. 177-192
- Kim, J. 1998: Philosophy of Mind (Westview Press: Boulder, Col.)
- Kripke, S.A. 1982: Wittgenstein on Rules and Private Language (Basil Blackwell: Oxford)
- 1980: Naming and Necessity (Harvard University Press: Cambridge, Mass.).
- Levine, J. 1983: 'Materialism and Qualia: The Explanatory Gap'; *Pacific Philosophical Quarterly*, vol. 64, no. 4, October, 1983, 354 - 361.
- Lewis, D. K. 1988. Desire as belief. *Mind* 97:323-32.
- 1986: On the Plurality of Worlds (Blackwell: New York)
- 1983: 'Mad Pain and Martian Pain'; Philosophical Papers I (OUP: New York).
- 1979: 'Attitudes "De Dicto" and "De Se"', *Philosophical Review*, O79, 88, pp. 513-543.

- 1973: 'Causation'; *Journal of philosophy* Vol. 70, pp. 556-567 [Reprinted in Lewis, D.K.: *Philosophical Papers II* (OUP: New York, 1975)]
1972. 'Psychophysical and theoretical identifications.' *Australasian Journal of Philosophy* 50:249-58. Reprinted in [N. Block, (ed)] *Readings in the Philosophy of Psychology* (MIT Press, 1980).
- 1970: 'How to define theoretical terms'; *Journal of Philosophy* Vol. 67, No. 13 (July, 1970)
- Locke, J. 1975: *An Essay Concerning Human Understanding* [Nidditch, P.H. (ed.)] (OUP: New York).
- Loos E.E. (ed) 2004: 'Glossary of Linguistic terms' (SIL Linguistics) accessed online 31/12/2005  
<http://ccat.sas.upenn.edu/~haroldfs/dravling/illocutionary.html>
- Lycan, W. G. 1971: 'Gombrich, Wittgenstein, and the Duck-Rabbit', *Journal of Aesthetics and Art Criticism*, Wint 71, 30, pp. 229-237
- Lyons, W. 1978: 'Emotions and Behaviour', *Philosophy and Phenomenological Research*

- Mackie, J. L. 1977: Ethics: Inventing Right and Wrong. (Penguin Books: London).
- Macpherson, F. 2005: 'Colour Inversion Problems for Representationalism', *Philosophy and Phenomenological Research* 70
- 2003: 'Novel Colours and the Content of Experience', *Pacific Philosophical Quarterly*, 84, pp. 43-66
- Martin, M. 2002: 'The Transparency of Experience', *Mind and Language*, 4, pp. 376-425
- Maund, B. 2002: 'Colour'; Stanford Encyclopedia of Philosophy Online (accessed 9/10/2005); <http://plato.stanford.edu/entries/color/#Subj>
- Maynard, P. 1994: 'Seeing Double', *Journal of Aesthetics and Art Criticism*, Spr 94, 52(2), pp. 155-167
- McDowell, J. 1994: Mind and World (OUP: New York)
- 1994: 'The Content of Perceptual Experience', *Philosophical Quarterly*, 44, pp. 190-205
- 1978: "Are Moral Requirements Hypothetical Imperatives?" *Proceedings of the Aristotelian Society*, 13-29.
- McGinn, C. 1983: The Subjective View: Secondary Qualities and Indexical Thoughts (Clarendon Press: New York).

- McNaughton, D. 1988: Moral Vision: An Introduction to Ethics (Basil Blackwell: New York).
- Mele, A. 2005: 'Decisions, Intentions and Free Will', *Midwest Studies in Philosophy*, 29, pp. 146-162
- 1999: 'Is There a Place for Intention in an Analysis of Intentional Action?', *Philosophia*, N 99, 27(3-4), pp. 410-432
- 1992: Springs of Action: Understanding Intentional Behaviour (OUP: New York)
- 1992: 'Intentions, Reasons, and Beliefs: Morals of the Toxin Puzzle', *Philosophical Studies*, N 92, 68(2), pp. 171-194
- Mellor, D. H. 1988: 'Crane's Waterfall Illusion', *Analysis*, Je 88, 48, pp. 147-150
- Mill, J.S. 1863: Utilitarianism [Crisp, R. (ed)] [OUP: New York (1998)].
- Millar, A. 2004: Understanding People: Normativity and Rationalising Explanation (Clarendon Press: Oxford)
- 1991: 'Concepts, Experience, and Inference', *Mind*, 100, pp. 495-505
- Millikan, R. 2004: Varieties of Meaning (MIT Press: Cambridge, MA)



- 1993: 'Content and Vehicle', in Spatial Representation  
[Eilan, N. (ed)], pp. 256-268 (Blackwell: Cambridge)
- 1991: 'Perceptual Content and Fregean Myth', *Mind*, O 91,  
100(399), pp. 439-459
- Moore, A. 2004: 'Hedonism'; Stanford Encyclopedia of Philosophy.  
Accessed Online (15/02/2006)  
<http://plato.stanford.edu/entries/hedonism/#1>
- Moore, G.E. 1903: Principia Ethica (Cambridge University Press:  
Cambridge)
- Morton, B. 1974: 'Beardsley's Conception of the Aesthetic Object',  
*Journal of Aesthetics and Art Criticism*, Spr 74, 32, pp. 385-  
396
- Nagel, T. 1970: The Possibility of Altruism (OUP: Oxford)
- Nash, R. A. 1989: 'Cognitive Theories of Emotion', *Nous*, 23, pp. 481-  
504
- Neu, J. 2000: A Tear is an Intellectual Thing: The Meanings of  
Emotion (OUP: New York)
- Nichols, S. & 2003: Mind-Reading (OUP: New York)
- Stich, S.
- Nida-Rumelin, M. 2002: 'Qualia: The Knowledge Argument' *Stanford  
Encyclopedia of Philosophy*

<http://plato.stanford.edu/entries/qualia-knowledge/#4.3>

(accessed 14/7/05)

- Noe, A. 2005: Action in Perception (MIT Press: Cambridge, Mass.)
- 1999: 'Thought and Experience.' *American Philosophical Quarterly*, July 1999, 36(3), pp. 257-265.
- Nussbaum, M. 2004: 'Emotions as Judgments of Value and Importance', in Thinking about Feeling: Contemporary Philosophers on Emotions [Solomon, R. (ed)] (OUP: New York)
- Oddie, G. 2005: Value, Reality and Desire (OUP: New York)
- O'Regan, K and Noe, A. 2001: 'What is it Like to See: A Sensorimotor Theory of Perceptual Experience', *Synthese*, 129, pp. 79-103
- Pacherie, E. 2002: 'The Role of Emotions in the Explanation of Action', *European Review of Philosophy*, 5, pp. 53-92
- 2000: 'Levels of Perceptual Content', *Philosophical Studies*, 100, 100(3), pp. 237-254
- 2000: 'The Content of Intentions', *Mind and Language*, 15, 15(4), pp. 400-432
- Peacocke, C. 1998: 'Nonconceptual Content Defended', *Philosophy and Phenomenological Research*, 58, 58(2), pp. 381-388

- 1992: 'Anchoring Conceptual Content: Scenarios and Perception' in Cognition, Semantics and Philosophy [Ezquerro, J. (ed)] (Kluwer: Norwell)
- 1992: 'Scenarios, Concepts and Perception' in The Contents of Experience [Crane, T. (ed)] (Cambridge Univ. Press: New York)
- 1983: Sense and Content: Experience, Thought, and their Relations (OUP: New York)
- Pendlebury, M. 1990: 'Sense Experiences and their Contents: A Defense of the Propositional Account', *Inquiry*, 33, pp. 215-30
- Perry, J. 1979: 'The Problem of the Essential Indexical', *Nous*, MR 79, 13, pp. 3-21.
- Platts, M. 1979: "Moral Reality." In Ways of Meaning: An Introduction to a Philosophy of Language (Routledge and Kegan Paul: London).
- Pollock, J. 1982: Language and Thought (Princeton University Press: Princeton).
- Price, H. H. 1964: 'Appearing and Appearances', *American Philosophical Quarterly*, Ja64, 1, pp. 3-19
- Prinz, J. J. 2005: 'Are Emotions Feelings?', *Journal of Consciousness Studies*, 12(8-10), pp. 9-25

- 2004: Gut Reactions: A Perceptual Theory of the Emotions  
(OUP: New York)
- 2004: 'Embodied Emotions', in Thinking about Feeling: Contemporary Philosophers on Emotions [Solomon, R. (ed)]  
(OUP: New York)
- Putnam, H. 1975: 'The Meaning of Meaning', Language, Mind and Knowledge. Minnesota Studies in the Philosophy of Science vol. 7, [Keith Gunderson (ed)] (Minneapolis: University of Minnesota Press), pp. 131-193. Reprinted in *Mind, Language and Reality* (1975), pp. 215-271.
- Quinn, W. 1989: 'Actions, Intentions, and Consequences: The Doctrine of Double Effect', *Philosophy and Public Affairs*, Fall 89, 19, pp. 334-351
- Railton, P. 1986: "Moral Realism." *Philosophical Review*, 95, pp. 163-207.
- Raz, J. 2001: Boundaries and Allegiances (OUP, New York).  
1999: Engaging Reason: On the Theory of Value and Action  
(OUP: Oxford)
- Roberts, R.C. 1996: 'Propositions and Animal Emotion', *Philosophy*, Jan 96, 71(275), pp. 147-156

- 1992: 'Emotions as Access to Religious Truths', *Faith and Philosophy*, Ja 92, 9(1), pp. 83-94
- Rosati, C.S. 2006: 'Moral Motivation': *Stanford Encyclopedia of Philosophy*, accessed online <http://plato.stanford.edu/entries/moral-motivation/> (17-02-2007)
- Russell, B. 1921: The Analysis of Mind (George Allen and Unwin: London)
- 1912: The Problems of Philosophy (Williams and Norgate: London)
- 1910: Principia Mathematica 3 Vols. (with A. N. Whitehead) (Cambridge University Press: Cambridge)
- 1903: The Principles of Mathematics (Cambridge University Press: Cambridge)
- Ryle, G. 1949. The Concept of Mind. (Hutchinson and Co.)
- Scanlon, T. M. 1998: What we Owe to Each Other (Harvard University Press: Harvard)
- 1982: 'Contractualism and Utilitarianism' in Utilitarianism and Beyond [Sen, A. and Williams, B. (eds)], pp. 103-128. (Cambridge University Press: Cambridge).

- Schleffer, I. 1991: In Praise of the Cognitive Emotions (Routledge: New York).
- Schroeder, T. 2004: Three Faces of Desire (OUP: New York)
- Schwitzgebel, E. 2002: 'A Phenomenal, Dispositional Account of Belief', *Nous*, 36, pp. 249-275
- Searle, J.R. 1983: Intentionality (Cambridge University Press: Cambridge)
- 1979: 'The Intentionality of Intention and Action', *Inquiry*, Autumn 79, 22, pp. 253-280
- 1969: Speech Acts (Cambridge University Press: Cambridge).
- Siegel, S. 2006: 'Which Properties are Represented in Perception?', in Perceptual Experience [Gendler, T. and Hawthorn, J. (ed)] (OUP: New York)
- 2004: 'Indiscriminability and the Phenomenal', *Philosophical Studies*, 120, pp. 91-112
- Shafer-Landau, R. 2003: Moral Realism: a defence (Clarendon: New York).
- Shoemaker, S. 1996: The First-Person Perspective and Other Essays, (Cambridge University Press, Cambridge).
- Skinner, B. F. 1953: Science and Human Behavior (McMillan: New York).

- Smart, J. J. C. 1959: 'Sensations and Brain Processes', *Philosophical Review*, Vol. 68, No. 2, pp. 141-156.
- Smith, A. D. 1999: The Problem of Perception (Harvard University Press: Cambridge, Mass.)
- Smith, M. 1998: 'Galen Strawson and the Weather Watchers.' *Philosophy and Phenomenological Research* Vol. 58 No. 2 (June 1998) pp. 449-454.
- 1994: The Moral Problem (Blackwell: Oxford)
- 1987: 'The Humean Theory of Motivation' *Mind* Vol.96, No.381. pp. 36-61.
- Sobel, D. and 2001: 'Against direction of fit accounts of belief and desire'
- Copp, D. *Analysis* 61.1, pp.44-53.
- Solomon, R. C. 1977: 'The Rationality of the Emotions', *Southwestern Journal of Philosophy*, Sum 77, 8, pp. 105-114
- Soltis, J. 1966: Seeing, Knowing and Believing: A Study of the Language of Visual Perception (Addison Wesley: Reading, MA)
- Sorensen, R. 2002: 'Meaningless Beliefs and Mate's Problem', *American Philosophical Quarterly*, Ap (02), 39(2), pp. 169-182.
- 1988: Blindspots (OUP: New York).
- Spinoza, B. 1677: The Ethics (Penguin Classics: London (2005))

- Stalnaker, R. 1984: Inquiry (MIT Press: Cambridge, Mass.)
- Stampe, D. 1987: 'The Authority of Desire'; *The Philosophical Review*, Vol. 96, No.3 pp. 335-381.
- Strawson, G. 1998: 'Précis of Mental Reality' & 'Reply to Smith'; *Philosophy and Phenomenological Research*, Vol. 58, No. 2 (June 1998) pp 433-454 & pp 471-483
- 1994: Mental Reality (OUP: New York)
- 1986: Freedom and Belief (OUP: New York).
- Stocker, M. 1979: 'Desiring the Bad; An Essay in Moral Psychology'; *Journal of Philosophy*, Vol. 76, No. 12 (Dec. 1979)
- 1970: 'Intentions and Act Evaluations', *Journal of Philosophy*, S 70, 67, pp. 589-602
- Stoljar, D. 2006: 'Two Conceivability Arguments Compared', *Proceedings of the Aristotelian Society*, Vol. CVII.
- 2001: 'Physicalism' *Stanford Encyclopedia of Philosophy*  
<http://plato.stanford.edu/entries/physicalism/> (accessed 14/7/05)
- Stone, T. and Young, A. 1997: 'Delusions and Brain Injury', *Mind and Language*, SD 97, 12(3-4), pp. 327-365.
- Svavarsdottir 1999: Reason, Emotion and Will (Ashgate Publishing Company: Aldershot).



- Tappolet, C. 2005: 'Ambivalent Emotions and the Perceptual Account of Emotions', *Analysis*, JI 05, 65:3(287), pp. 229-233
- Thau, M. 2004: 'What is Disjunctivism?', *Philosophical Studies*, JI-Ag-S 04, 120(1-3), pp. 193-253
- Tienson, J. L. 1984: 'Hesperus and Phosphorus', *Australasian Journal of Philosophy*, MR 84; 62. pp. 16-25.
- Tye, M. 2002: 'Representationalism and the Transparency of Experience', *Nous*, Vol. 36, 1, pp. 137-151.
- 2000: Consciousness, Colour, and Content (MIT Press: Cambridge, Mass.)
- 1995: Ten Problems of Consciousness (MIT Press: Cambridge, Mass.)
- 1994: 'Qualia, Content and the Inverted Spectrum', *Nous*, Je 94, 28(2), pp. 159-183
- Valberg, J. J. 1992: The Puzzle of Experience (OUP: New York)
- Van Gulick, R. 1994: 'Are Beliefs Brain States? And if they are what might that explain?' *Philosophical Studies*, 76, pp. 205-212
- Van Inwagen, P. 2003: 'Existence, Ontological Commitment, and Fictional Entities.' in The Oxford Handbook of Metaphysics [Loux, M.J. (ed.)] (OUP: Oxford).

- Van Roojen, M. 2002: 'Should Motivational Humeans be Humeans about Rationality?': *Topoi*. 21.
- Velleman, J.D. 2000: 'On the Aim of Belief', Essay 11; The Possibility of Practical Reason (OUP: New York)
- 1992: 'The Guise of the Good'; *NOUS* 26:1, pp. 3-26
- 1992: 'What Happens When Someone Acts?' *Mind* Vol. 101, No. 403 (July 1992) pp. 461-481.
- 1989: Practical Reflection (Princeton University Press: Princeton)
- Wallace, R.J. 2003: 'Practical Reason'; *Stanford Encyclopedia of Philosophy* (<http://plato.stanford.edu/entries/practical-reason/>)
- Walton, K.L. 1990: Mimesis as Make-Believe (Harvard University Press: Cambridge, Mass.)
- Weintraub, R. 2001: 'The Lottery: A Paradox Regained and Resolved', *Synthese*, D01, 129(3), pp. 439-449.
- Wheeler, G. R. 2005: 'On the Structure of Rational Acceptance: Comments on Hawthorne and Bovens', *Synthese*, Mr II 2005, 144(2), pp. 287-304.
- Wiggins, D. 1987: Needs, Values and Truth: Essays in the Philosophy of Value (OUP: Oxford)

- Williams, B. 1973: Problems of the Self (Cambridge University Press: Cambridge)
- Williamson, T. 1996: 'Knowing and Asserting', *Philosophical Review*, O96, 105(4), pp. 489-523.
- Wolheim, R. 1973: On Art and the Mind (Allen Lane: London)
- Wright, C. 1989: 'Wittgenstein's Rule-following Considerations and the Central Project of Theoretical Linguistics'; Reflections on Chomsky [George, A. (ed)] (Basil Blackwell: New York).
- Zangwill, N. 1998: 'Direction of fit and normative functionalism.' *Philosophical Studies* 91, pp.173-203.
- Ziff, P. 1951: 'Art and the "Object of Art"', *Mind*, O51, 60, pp. 466-480